Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Light-Weight Semantic Segmentation Based on Mask RCNN

Mingwei Liu *

School of Computer Science, University of Dublin, Ireland

* Corresponding Author Email: mingwei.liu@ucdconnect.ie

Abstract. Semantic segmentation is based on the image information to detect and identify various categories of objects and output these objects mask. Compared with object detection and image classification, semantic segmentation has a more accurate recognition effect, and has a wide range of applications in automatic driving and other fields. With the development of deep learning, semantic segmentation methods based on deep learning have achieved good results, such as mask rcnn. Nowadays, Mask RCNN already has a good ability to handle object segmentation and mask the object. However, the model needs to consume a lot of computational power so that raise request for the equipment. This paper proposes a lightweight semantic segmentation network based on Mask RCNN, which can achieve better semantic segmentation accuracy with less data. In particular, using deep convolution as the feature extraction operator can effectively reduce the calculation and parameter number of the model. Experiments show the effectiveness of the proposed method.

Keywords: Convolutional Neural Network; Depth-wise Convolution; Lightweight Neural Network; Semantic segmentation.

1. Introduction

Image semantic segmentation is a basic task in the field of image analysis, which has been widely used in pedestrian detection, automatic driving and industrial defect detection. Unlike image classification and object detection, semantic segmentation aims to assign a class label to each pixel in an image, providing a more detailed and precise description of the image. With the development of machine learning, the end-to-end training method based on deep learning makes the semantic segmentation algorithm reach a better level. However, deploying these algorithms directly on end devices for inference poses challenges due to limited computational resources. One such algorithm, Mask RCNN, has achieved high accuracy and speed in semantic and instance segmentation tasks. Even though, conventional Mask RCNN models raises significant computational demands, which pose challenges for their deployment on lightweight mobile devices [1]. Typically, semantic segmentation methods based on deep learning utilize Convolutional Neural Networks (CNNs) as feature extraction kernels and are trained end-to-end to achieve optimal performance [2]. Nevertheless, traditional CNNs employing standard convolutional kernels contain a large number of parameters, necessitating abundant data for effective model fitting, thereby substantially increasing training costs [2]. Moreover, CNN-based approaches suffer from slow computation speeds and heavy reliance on computational resources [3], rendering them impractical for inference on devices such as smartphones and cameras, consequently severely restricting their real-world application potential.

In this paper, Feather Mask RCNN semantic segmentation model is proposed based on Mask RCNN, which can improve the detection speed while maintaining high segmentation accuracy [4]. Specifically, the model consists of a backbone network for feature extraction and a lightweight semantic segmentation module for mask prediction. Use the depth wise convolutional module to replace the common convolution in the backbone and semantic segmentation module. It can effectively reduce the number of parameters in the model and improve the detection speed.

2. Related Work

2.1. Mask RCNN

Mask RCNN is composed of feature extraction module, object detection module, classification module and semantic segmentation module as shown in Fig. 1. In the case of semantic feature extraction, the proposed RoI-Align can achieve more refined feature and position information extraction. Optimal results were achieved at the time on a variety of data sets [1].

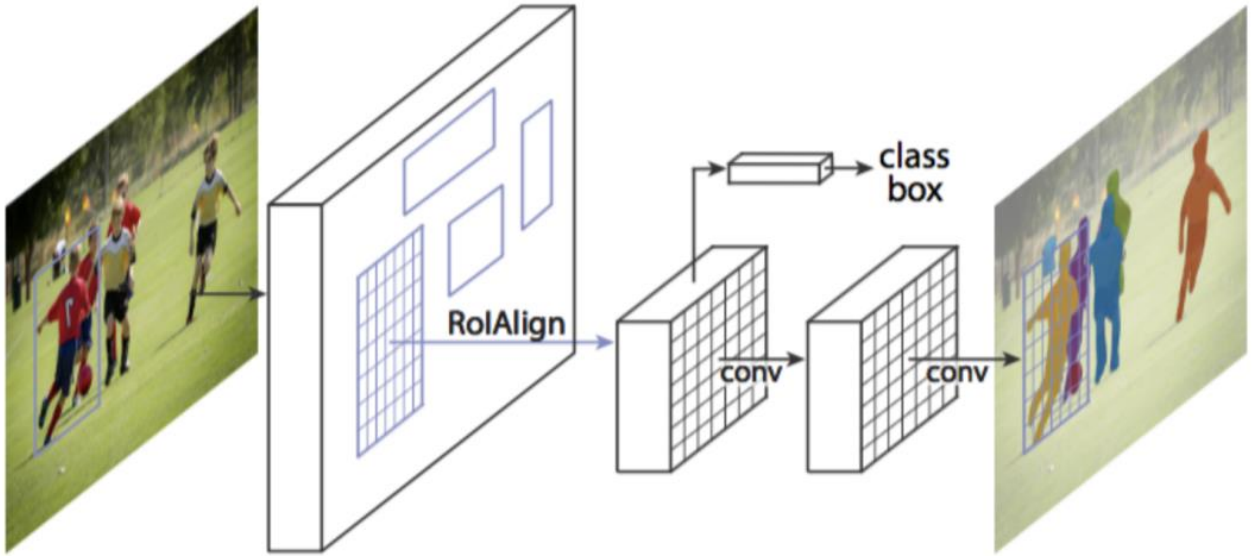

Fig. 1 Mask RCNN [1]

2.2. Depth-wise Convolution

Depth wise Convolution is a lightweight convolution operation commonly used in artificial intelligence neural networks. It extracts the input features by depth wise convolution, and then changes the dimensional by point convolution.

In depth wise convolution, each input channel is convolved independently with its own set of filters. This means that each channel is processed separately, resulting in a set of feature maps. After depth wise convolution, pointwise convolution is applied, which performs a 1x1 convolution over all channels of the feature maps obtained from depth wise convolution as shown in Fig. 2 [5]. The goal of this step is to combine the spatial information captured by the depth wise convolution across different channels [5].

The main advantage of depth-wise convolution is that it achieves feature processing similar to standard convolution with a small number of parameters [6]. Because each input channel has its own set of filters, the number of parameters required is much lower than in traditional convolutions, where a single set of filters is applied to all channels [6]. This reduction in parameters not only results in smaller models, but also reduces computational costs, making depth wise convolution particularly suitable for devices with low computing power and storage, such as cameras, mobile phones, wearable devices, and so on.

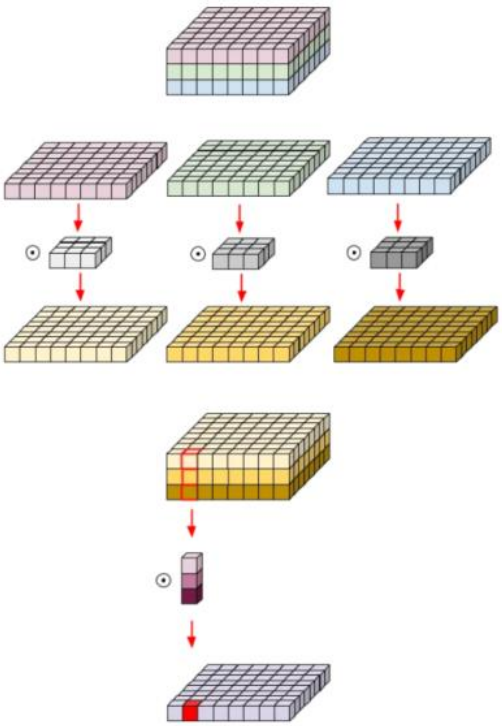

Fig. 2 Depth-wise convolution rationale [7]

2.3. Backbone

The backbone of the network is used as a feature extraction model in image classification, object detection, semantic segmentation and other tasks. Classic backbone networks include VGG, ResNet, swin transformer, ViT, DenseNet and so on [8]. The selection of backbone network has a great impact on the effect and performance of the model, and in practical applications, there is often a trade-off between speed and accuracy.

3. Mehtod

3.1. Feather Mask RCNN

Conventional convolution operations rely on a large amount of computing power and are slow to reason on devices with lower computing resources. Feather RCNN proposed in this paper is to solve this problem. Feather RCNN is composed of feature extraction module, target detection module and semantic segmentation module [9, 10]. The image goes through the feature extraction module to obtain the deep semantic information of the image, the object detection module predicts the category and position of the target, and finally the semantic features of the target are extracted through RoI Align. As shown in Fig. 3.

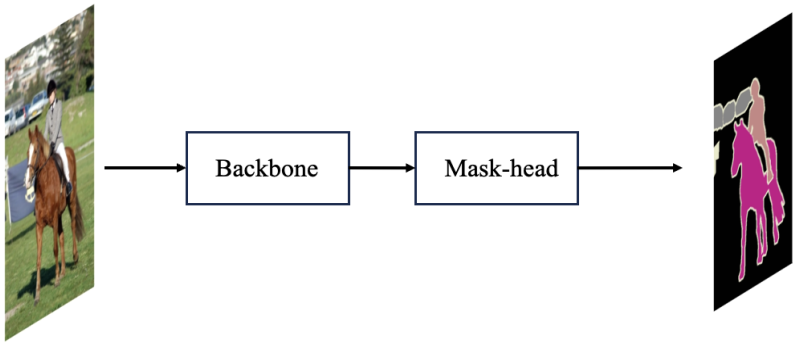

Fig. 3 Structure of Feather Mask RCNN

3.2. Experiment

3.2.1. Dataset

This paper chooses the same training set COCO as Mask-RCNN for training. COCO dataset is rich in training resources to help our network train and achieve the desired accuracy. COCO is an open-source atlas for the Open Challenge for Image Recognition, which contains 330k images of daily life and 800k categories of common objects that we will encounter in our life such as fruits, chairs, etc. These data can help feather mask rcnn learn efficiently in instance segmentation and semantic understanding, as shown in Fig. 4.

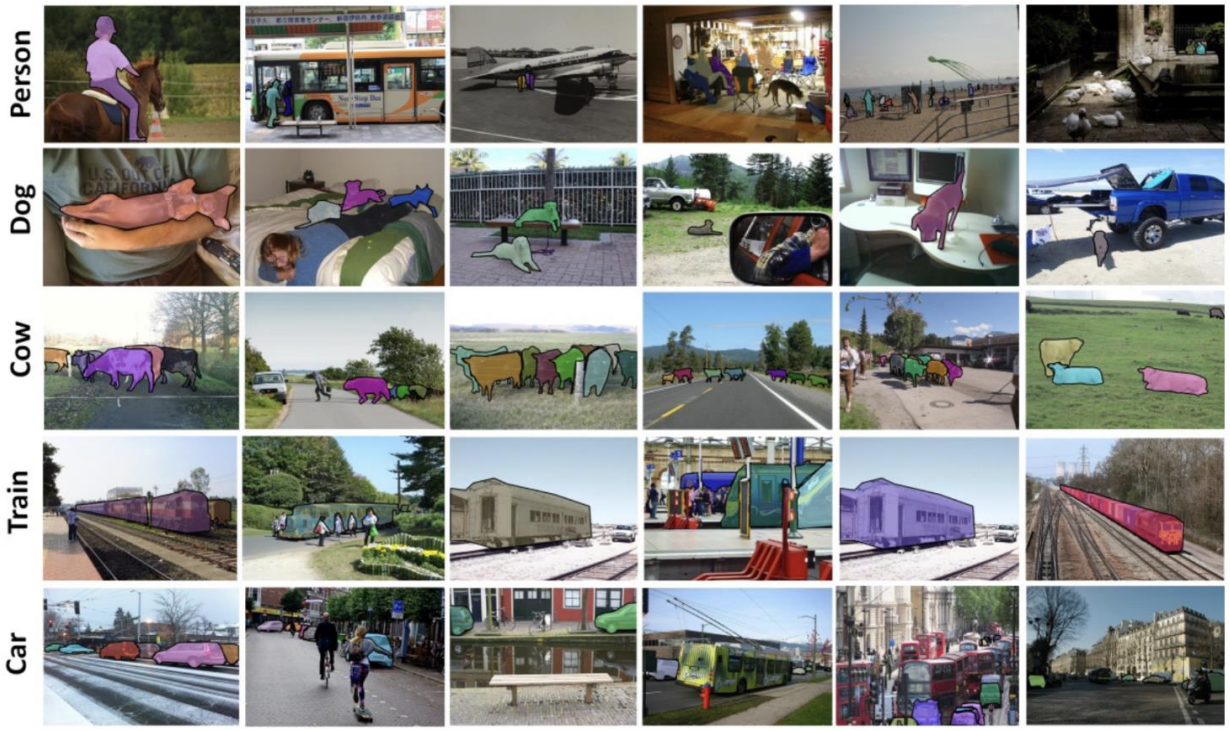

Fig. 4 Samples of the COCO dataset [11]

3.2.2. Parameter setting

The experiment used Resnet50 as the backbone network, and replaced all the convolutional operations with Depth wise. In the experiment, SGD was used as the optimizer, the learning rate was set to 0.002 and the batch size was set to 8. The model converges after 20 epochs are trained on the COCO dataset.

3.3. Result Analysis

The experimental results show that using deep wise convolution can reduce the computational amount of the model, making it possible to run on devices with lower computational power. As the number of parameters in the model decreases, the model becomes more difficult to learn and requires longer training time to achieve fitting.

4. Conclusion

The paper raises a light version RCNN called feather RCNN which solves a lot of issues of the state-of-the-art of neutral network, especially reducing the usage complexity, which makes more efficient and smarter mobile applications feasible. This not only improves the operational efficiency of mobile devices and reduces energy consumption, but also optimizes the user experience and expands the application scenarios of mobile applications. Neural network light-weighting techniques can help compress models to a size that fits the resource constraints of mobile devices and enable real-time image recognition on mobile devices.

In addition to this, neural network light-weighting technology can also help deploy efficient deep learning models in cars for recognizing pedestrians, vehicles, traffic signs, etc., helping to solve the problem of autonomous driving technology that requires high computing power support. In summary, lightweight networks is of great significance on mobile devices, providing new opportunities and possibilities to promote the application and development of deep learning technology on mobile. More efficient compression of models, and accuracy in maintaining changes while efficiently compressing models is the topic that can be further studied.

References

[1] K. He, G. Gkioxari, P. Dollár and R. Girshick, Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV), 2017, 2980-2988.

[2] Park, Jinyoung, and Hoseok Moon. Lightweight Mask RCNN for Warship Detection and Segmentation. IEEE Access, 2022, 10 : 24936–24944.

[3] Sun, Ling, et al. LAD-RCNN: A Powerful Tool for Livestock Face Detection and Normalization. Animals (Basel), 2023, 13(9) : 1446.

[4] Wang, Xi, et al. multi-scale coal and gangue detection in dense state based on improved Mask RCNN. Measurement: Journal of the International Measurement Confederation, 2023, 221: 113467.

[5] Mubarak, Auwalu Saleh, et al. Effect of Gaussian filtered images on Mask RCNN in detection and segmentation of potholes in smart cities. Mathematical Biosciences and Engineering : MBE, 2023, 20(1) : 283–295.

[6] Amudhan, A. N., and A. P. Sudheer. Lightweight and computationally faster Hypermetropic Convolutional Neural Network for small size object detection. Image and Vision Computing, 2022, 119: 104396.

[7] Atul Pandey, Depth-wise Convolution and Depth-wise Separable Convolution， 2018.

[8] Zhang, Wenchao, et al. Global context aware RCNN for object detection. Neural Computing & Applications, 2021, 33(18) : 11627–11639.

[9] Fatima, Anum, et al. Deep Learning-Based Multiclass Instance Segmentation for Dental Lesion Detection. Healthcare (Basel), 2023, 11(3) : 347.

[10] Jia, Weikuan, et al. Accurate segmentation of green fruit based on optimized mask RCNN application in complex orchard. Frontiers in Plant Science, 2022, 13 : 955256–955256.

[11] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context, Computer Vision–ECCV 2014: 13th European Conference, 2014, 13 : 740-755.