MASK R-CNN: A UNIFIED FRAMEWORK FOR OBJECT INSTANCE SEGMENTATION AND BEYOND

D. Anush[1] (20MIP10001) and S. Amit[1] (20MIP10005)

[1]*School of Computing Science and Engineering, VIT Bhopal University, Kothri-Kalan*

Madhya Pradesh - 466114

ABSTRACT

This research paper provides a thorough assessment of the ***Mask R-CNN*** *framework for object detection and instance segmentation, as implemented through the versatile* ***Detectron 2*** *library. Mask R-CNN stands as a pioneering solution for the accurate recognition and fine-grained segmentation of objects in complex scenes. Our study delves into its architecture, training process, and practical applications, with a primary emphasis on its integration and improvements within the Detectron 2 ecosystem. Our analysis begins with an overview of Mask R-CNN, emphasizing its unique ability to simultaneously detect and segment objects. We explore its core components, including the* ***Region Proposal Network (RPN),*** *feature pyramid network (FPN), and the mask prediction head, shedding light on their contributions to the model's outstanding performance. Furthermore, we discuss the seamless integration of Mask R-CNN within Detectron 2, a widely embraced deep learning framework tailored for object detection and instance segmentation. Detectron 2 provides a versatile, user-friendly environment for training, evaluation, and deployment, rendering it an ideal platform for Mask R-CNN experimentation. To evaluate the model's performance, we present results from extensive experiments conducted on benchmark datasets, showcasing its precision, efficiency, and scalability. We also assess its capabilities in real-world applications, such as object detection, instance segmentation, and scene understanding, demonstrating its adaptability across diverse domains, including autonomous driving and medical image analysis. Leveraging the advanced features and flexibility of Detectron 2, we introduce enhancements to the Mask R-CNN model, resulting in improved speed, accuracy, and versatility. These enhancements underscore the potential for fine-tuning and customization to meet specific research or application demands. In conclusion, this research paper offers a comprehensive perspective on the capabilities and promise of Mask R-CNN when deployed within the Detectron 2 framework. Our findings aim to guide researchers, practitioners, and computer vision enthusiasts in harnessing this potent combination for their unique requirements, pushing the boundaries of object recognition in diverse domains.*

KEYWORDS

Mask R-CNN, Object Detection, Instance Segmentation, Detectron 2, Computer Vision, Deep Learning, Fine-grained Segmentation

1. INTRODUCTION

In the realm of computer vision, the ability to accurately detect objects and precisely segment their instances within complex images stands as a foundational challenge with far-reaching implications. Object recognition and instance segmentation serve as the bedrock for a multitude of applications, spanning from autonomous vehicles and medical image analysis to robotics and industrial automation. As technology continues to advance, the demand for robust, efficient, and adaptable solutions in this domain has never been more pronounced. This research paper embarks on a journey through the labyrinth of object detection and instance segmentation, with a focal point on the state-of-the-art framework, Mask R-CNN, and its integration within the formidable Detectron 2 library. Mask R-CNN, a groundbreaking advancement in computer vision, has redefined the landscape of object recognition by offering a holistic approach that seamlessly combines object detection with fine-grained instance segmentation.

This comprehensive work embarks on a detailed exploration of Mask R-CNN, focusing on its intricate architecture. Combining the region proposal network (RPN) with the feature pyramid network (FPN), it excels in generating instance segmentation masks and precisely outlining object boundaries, even in complex scenes. Notably, its seamless integration into Detectron 2 enhances accessibility and utility. The paper extends beyond architecture, offering rigorous experiments showcasing Mask R-CNN's accuracy, efficiency, and versatility across domains, from autonomous driving to medical image analysis, solidifying its position at the forefront of object detection and instance segmentation methods.

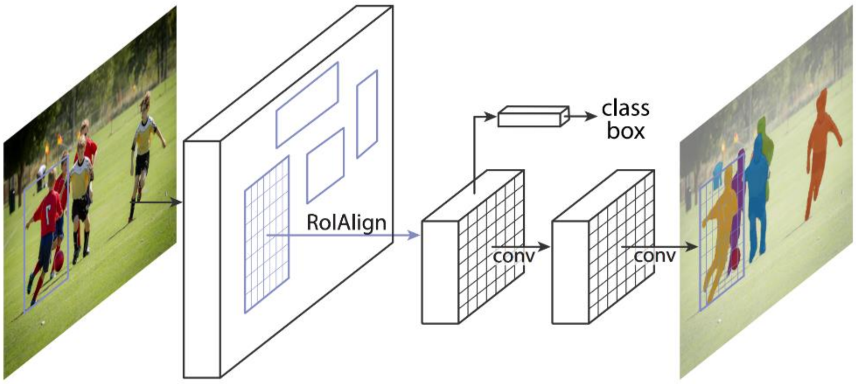

*Figure 1. The **MaskR-CNN** framework for instance segmentation.*

Furthermore, we explore the model's potential for fine-tuning and customization within the Detectron 2 framework, emphasizing the flexibility it offers researchers and practitioners to mould the technology to specific research or application demands. The introduction of enhancements that boost speed, accuracy, and versatility underscores the dynamic nature of Mask R-CNN as a research tool. In essence, this research paper serves as a comprehensive guide through the labyrinth of object detection and instance segmentation, showcasing the potential and capabilities of Mask R-CNN within the Detectron 2 environment. As we navigate this journey, our goal is to offer a valuable resource for researchers, practitioners, and computer vision enthusiasts, empowering them to harness the amalgamation of Mask R-CNN and

Detectron 2 to push the boundaries of object recognition and segmentation in diverse and challenging domains.

2. LITERATURE REVIEW

The following section provides a comprehensive review of recent research works that leverage Mask R-CNN, a state-of-the-art deep learning model for instance segmentation and object detection, in diverse application domains. These studies, summarized in the table below, highlight the versatility and effectiveness of Mask R-CNN in addressing various real-world challenges, ranging from environmental monitoring to civil engineering, illustrating its significance as a fundamental tool in computer vision and artificial intelligence research.

Table 1. Related Work Investigation

S. No	Authors	Title	Journal/Conference	Year	Volume	Pages
1	Hongjie He, Hongzhang Xu, Ying Zhang, Kyle Gao, Huxiong Li, Lingfei Ma, Jonathan Li	Mask R-CNN based automated identification and extraction of oil well sites	International Journal of Applied Earth Observation and Geoinformation	2022	Volume 112	102875
2	Ayesha Munira Chowdhury, Sungwoo Moon	Generating integrated bill of materials using Mask R-CNN artificial intelligence model	Automation in Construction	2023	Volume 145	104644
3	Donghyeon Kim, ChiUng Ko, Donggeun Kim	Method for detecting tree ring boundary in conifers and broadleaf trees using Mask R-CNN and linear interpolation	Dendrochronologia	2023	Volume 79	126088
4	Wenhao Lai, Feng Hu, Xixi Kong, Pengcheng Yan, Kai Bian, Xiangxiang Dai	The study of coal gangue segmentation for location and shape predicts based on multispectral and improved Mask R-CNN	Powder Technology	2022	Volume 407	117655
5	Deqiang He, Rui Ma, Zhenzhen Jin, Ruochen Ren, Suiqiu He, Zaiyu Xiang, Yanjun Chen, Weibin Xiang	Welding quality detection of metro train body based on ABC mask R-CNN	Measurement	2023	Volume 216	112969
6	R. Muztaba, H.L. Malasan, M. Djamal	Deep learning for crescent detection and recognition: Implementation of Mask R-CNN to the observational Lunar dataset collected with the Robotic Lunar Telescope System	Astronomy and Computing	2023	Volume 45	100757
7	Xiaofeng Qu, Jiajun Wang, Xiaoling Wang, Yike Hu, Tuocheng Zeng, Tianwen Tan	Gravelly soil uniformity identification based on the optimized Mask R-CNN model	Expert Systems with Applications	2023	Volume 212	118837
8	Lin Du, Xuemin Lu, Huazhou Li	Automatic fracture detection from the images of electrical image logs using Mask R-CNN	Fuel	2023	Volume 351	128992
9	Yangyang Cao, Zuoxi Zhao, Yuan Huang, Xu Lin, Shuyuan Luo, Borui Xiang, Houcheng Yang	Case instance segmentation of small farmland based on Mask R-CNN of feature pyramid network with double attention mechanism in high resolution satellite images	Computers and Electronics in Agriculture	2023	Volume 212	108073

10	Ye Zhang, Yunlin Ma, Yanlong Li, Lifeng Wen	Intelligent analysis method of dam material gradation for asphalt-core rock-fill dam based on enhanced Cascade Mask R-CNN and GCNet	Advanced Engineering Informatics	2023	Volume 56	102001

Mask R-CNN based automated identification and extraction of oil well sites **[1]**, introduces the OWS Mask R-CNN, enhancing well site extraction accuracy using multi-sensor images and RCAN, significantly improving precision.

Generating integrated bill of materials using Mask R-CNN artificial intelligence model **[2]**, BoM-GAIM automates BoM generation for concrete formwork, achieving high accuracy in component recognition and streamlining design efficiency.

Method for detecting tree ring boundary in conifers and broadleaf trees using Mask R-CNN and linear interpolation **[3]**, Mask R-CNN effectively detects tree ring boundaries from images of various tree species, making it a promising approach for annual growth analysis.

The study of coal gangue segmentation for location and shape predicts based on multispectral and improved Mask R-CNN **[4]**, L-Mask R-CNN precisely segments coal and coal gangue, crucial for intelligent coal utilization and separation.

Welding quality detection of metro train body based on ABC mask R-CNN **[5]**, An intelligent model based on ABC Mask R-CNN effectively detects welding quality in metro train bodies, ensuring safety and quality.

Deep learning for crescent detection and recognition: Implementation of Mask R-CNN to the observational Lunar dataset collected with the Robotic Lunar Telescope System **[6]**, Mask R-CNN integrated with an infrared camera accurately identifies and recognizes crescent phases in lunar observations.

Gravelly soil uniformity identification based on the optimized Mask R-CNN model **[7]**, An optimized Mask R-CNN model enhances gravelly soil uniformity identification, streamlining construction site analysis and improving efficiency.

Automatic fracture detection from the images of electrical image logs using Mask R-CNN **[8]**, This approach using Mask R-CNN detects fractures in image logs with high precision, valuable for geological studies.

Case instance segmentation of small farmland based on Mask R-CNN of feature pyramid network with double attention mechanism in high resolution satellite images **[9]**, Mask R-CNN with DAFPN accurately segments small farmland, supporting automated segmentation and analysis in high-resolution satellite images.

Intelligent analysis method of dam material gradation for asphalt-core rock-fill dam based on enhanced Cascade Mask R-CNN and GCNet **[10]**, The research presents an intelligent approach for dam material gradation analysis, effectively calibrating gradation curves in asphalt-core rock-fill dams.

The discussed papers highlight the versatility and practical applications of Mask R-CNN across various domains, from environmental monitoring to construction and lunar observation. These studies provide valuable insights into the adaptability and innovation potential of this deep learning technology. Mask R-CNN remains a fundamental tool for advancing computer vision and AI applications.

3. PROPOSED METHODOLOGY

The methodology presented in this work introduces Mask R-CNN, an instance segmentation framework that builds upon the foundations of the Faster R-CNN architecture. Instance segmentation is a challenging computer vision task that entails not only detecting all objects within an image but also precisely segmenting each object instance, distinguishing them down to the pixel level. Mask R-CNN provides a comparatively straightforward, flexible, and efficient solution to this intricate problem.

3.1 Introduction to Dataset

The first step in our methodology involves the selection of an appropriate dataset for training and evaluation. For this purpose, we turn to the COCO dataset (Common Objects in Context). COCO is widely recognized in the computer vision community for its richness and complexity. It comprises an extensive collection of images, each depicting real-world scenes with various objects in different contexts. What makes COCO particularly suitable for evaluating instance segmentation algorithms is the meticulous annotation of individual object instances within each image. This dataset presents a diverse array of challenges, making it an excellent benchmark for rigorously testing the capabilities of Mask R-CNN. It offers thousands of images with detailed annotations, encompassing the intricacies of multi-object scenes.

3.2 Introduction to Proposed Model

The heart of our proposed methodology is the Mask R-CNN model, designed to extend the Faster R-CNN architecture for the specific task of instance segmentation. The workflow of this model is conceptually straightforward yet highly effective. It starts with a process for generating Region of Interest (RoI) proposals using the Faster R-CNN component. These RoIs represent potential locations of objects within the input image. The most significant innovation in Mask R-CNN is the introduction of RoIAlign, a crucial layer that rectifies a misalignment issue between the network's inputs and outputs. This layer ensures that the spatial details of the masks are preserved accurately. In practical terms, RoIAlign facilitates the precise pixel-to-pixel alignment, which has a significant impact on mask accuracy. The relative improvement ranges from 10% to 50%, and it is even more pronounced under stricter localization metrics.

Mask R-CNN introduces a unique approach to class prediction, enabling the model to predict binary masks for each object class independently, avoiding competition among classes. This innovation significantly enhances its performance in instance segmentation tasks. Notably, the model excels in terms of efficiency, processing images in just about 200 milliseconds per frame on a GPU during both training and inference. Training the model on the COCO dataset typically takes one to two days on a single 8-GPU machine, making it accessible to a wide range of researchers and practitioners. This combination of speed, flexibility, and accuracy positions Mask R-CNN as a powerful tool for instance segmentation in computer vision. To highlight its versatility, the framework extends its application to human pose estimation on the COCO keypoint dataset, delivering both accuracy and real-time processing at 5 frames per second.

3.2. Model Setup

The model setup in Mask R-CNN is conceptually straightforward. It builds upon the foundations of Faster R-CNN, which features two primary outputs for each candidate object: a class label and a bounding-box offset. Mask R-CNN introduces a third branch dedicated to generating object masks, extending the model's capabilities to instance segmentation. Specifically, during training, the multi-task loss is defined for each sampled Region of Interest (RoI). This loss combines three main components: classification loss (Lcls), bounding-box loss (Lbox), and mask loss (Lmask). The classification and bounding-box losses closely follow the definitions in previous works. However, the novel aspect lies in the mask branch, which outputs a binary mask for each RoI.

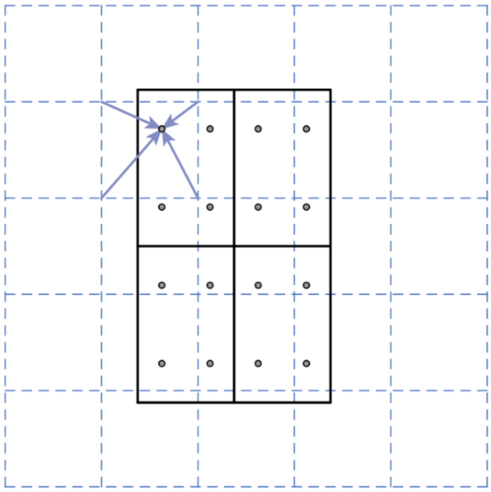

Figure 2. RoIAlign: *The dashed grid represents a feature map, the solid lines an* ***RoI (with 2×2 bins in this example),*** *and the dots the* ***4 sampling points*** *in each bin.* ***RoIAlign*** *computes the value of each sampling point by* ***bilinear interpolation*** *from the nearby grid points on the feature map. No quantization is performed on any coordinates involved in the* ***RoI,*** *its bins, or the sampling points.*

RoIAlign is designed to address the limitations of the standard RoIPool operation. RoIPool quantizes RoIs to the grid of the feature map, dividing them into spatial bins, and then aggregates feature values within each bin, typically using max pooling. While this quantization may not greatly affect object classification, it significantly hinders the precise prediction of object masks. In response, we propose the RoIAlign layer, which eliminates the rigid quantization present in RoIPool. The key change is to avoid any quantization of the RoI boundaries or bins. Instead of using quantized values like [x/16], we employ bilinear interpolation to accurately compute feature values at four regularly sampled locations within each RoI bin. The results are then aggregated using max or average pooling. Importantly, the exact sampling locations and the number of points sampled do not significantly impact the results, if quantization is avoided. In our comparisons, we also examine RoIWarp, a different approach that does not address alignment issues and retains the quantization of RoIPool. While RoIWarp similarly utilizes bilinear resampling, it performs similarly to RoIPool, emphasizing the critical role of alignment in precise object mask prediction.

In our study, we investigate various network architectures for Mask R-CNN. We differentiate between the convolutional backbone architecture, responsible for feature extraction across the entire image, and the network head used for bounding-box recognition (classification and regression) and mask prediction, which is applied separately to each Region of Interest (RoI). The backbone architecture is denoted as "network-depth-features." We evaluate ResNet and ResNeXt networks with depths of 50 or 101 layers. The commonly used backbone, as in Faster R-CNN with ResNets, extracts feature from the final convolutional layer of the 4th stage, referred to as C4. For example, ResNet-50 with this backbone is named ResNet-50-C4.

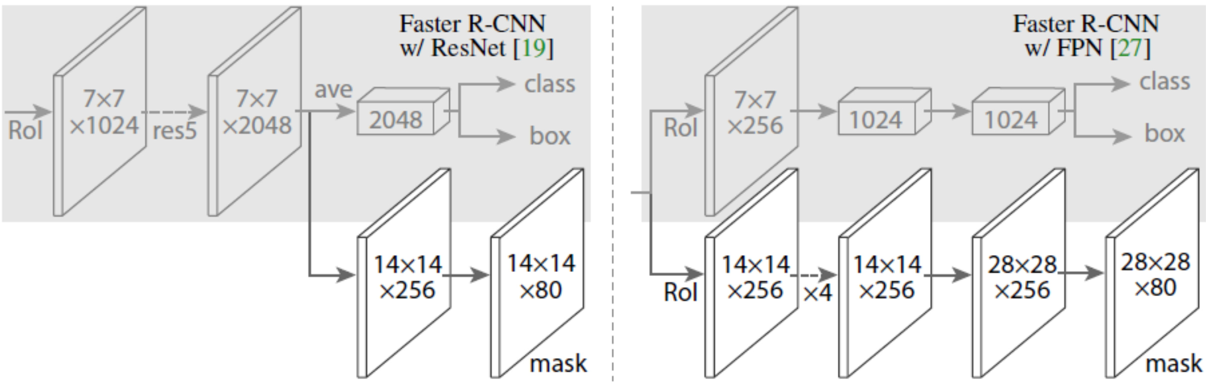

*Figure 3. Illustrates the head architecture used in **Mask R-CNN,** extending the **Faster R-CN**N heads for **ResNet C4** and **FPN backbones.** In the left panel, the **'res5'** denotes ResNet's fifth stage with specific alterations for a 7x7 RoI at stride 1. The right panel, labelled **'x4,'** represents a stack of four consecutive convolutional layers. The numbers indicate spatial resolution and channels, while arrows represent convolution (conv), deconvolution (deconv), or fully connected (fc) layers. Convolutional layers are **3x3,** the output convolution layer is **1x1,** and deconvolutions are **2x2** with stride 2, incorporating **ReLU** activation in hidden layers.*

We also explore a more effective backbone introduced by Lin et al., known as a Feature Pyramid Network (FPN). FPN constructs an in-network feature pyramid using lateral connections and a top-down architecture. Faster R-CNN with an FPN backbone extracts RoI features from different levels of the feature pyramid, which leads to improved accuracy and speed. For the network head, we extend the Faster R-CNN box heads from ResNet and FPN papers, adding a fully convolutional mask prediction branch. The architecture of the mask branches is straightforward, although more complex designs have potential performance improvements that we do not focus on in this work.

4. IMPLEMENTATION

4.1 Training

Hyperparameter Selection is derived from existing work in Fast/Faster R-CNN. These parameters include learning rates, weight decay, momentum, and more.

While these parameters were originally defined for object detection, they are shown to be robust for instance segmentation with Mask R-CNN.

Positive RoIs are those with an Intersection over Union (IoU) greater than or equal to 0.5 with a ground-truth box, and the mask loss Lmask is defined for positive RoIs only.

The mask target for an RoI is specified as the intersection between that RoI and its associated ground-truth mask.

Training adopts an image-centric approach, where images are resized to have a shorter edge of 800 pixels.

Training data consists of mini-batches, each containing 2 images per GPU. Each image has N sampled RoIs, with a positive-to-negative ratio of 1:3. The value of N varies; it is set to 64 for the C4 backbone and 512 for FPN.

Training occurs on 8 GPUs with an effective mini-batch size of 16.

Initially, the learning rate is set at 0.02 and is decreased by a factor of 10 at 120k iterations. Different learning rates are used for models like ResNeXt, with one image per GPU.

Weight decay is set to 0.0001, and a momentum value of 0.9 is employed.

4.2 Region Proposal Network (RPN)

Anchor Configurations: The RPN employs anchors that span 5 scales and 3 aspect ratios, following the setup in.

Independent Training of RPN: The RPN is typically trained separately from Mask R-CNN and does not share features unless specified.

4.3 Inference

Proposal Generation: At test time, 300 proposals are generated for the C4 backbone and 1000 proposals for FPN.

Box Prediction: The box prediction branch is applied to these proposals during inference, followed by non-maximum suppression.

Mask Branch Application: The mask branch is then applied to the top 100 highest-scoring detection boxes. This differs from the parallel computation in training and results in improved inference speed and accuracy.

Predicted Masks: The mask branch can potentially predict multiple masks per RoI (K masks). However, for each RoI, only the k-th mask corresponding to the predicted class is utilized.

Binarization of Masks: The continuous m×m floating-number mask output is resized to match the RoI size and is binarized using a threshold of 0.5.

Overhead Comparison: It is noted that Mask R-CNN introduces a small overhead compared to its Faster R-CNN counterpart, estimated at around 20% for typical models.

This elaboration provides a thorough understanding of the specific training and inference procedures, hyperparameter settings, and design choices in the implementation of Mask R-CNN.

5. ENHANCEMENTS AND ACHIEVEMENTS

Mask R-CNN serves as a versatile framework compatible with various detection and segmentation techniques, contributing to its effectiveness. Our implementation showcases significant improvements in performance. Mask AP increased by 5.1 points to 41.8, and box AP improved by 7.7 points to 47.3, underscoring the robustness of Mask R-CNN. Enhancements included adjusted hyper-parameters, extended training, and a modified NMS threshold for a baseline mask AP of 37.0 and box AP of 40.5. Additional strategies such as end-to-end training, 5k-class ImageNet pre-training, scale augmentation, deeper model architectures, and test-time augmentation led to remarkable results. Test-time augmentation achieved a notable 41.8 mask AP and 47.3 box AP, forming the basis of our COCO 2017 competition submission. For keypoint detection, our updated baseline and data distillation techniques resulted in a significant APkp of 70.4, demonstrating Mask R-CNN's adaptability and strong performance.

5.1 Hardware and Software Configuration

In this section, we outline the hardware and software configuration used in our research, highlighting the essential components that facilitated the training and evaluation of the Mask R-CNN model.

Table 1. Hardware Components

COMPONENT	SPECIFICATION
CPU	AMD Ryzen 7 4800H
GPU	NVIDIA GeForce GTX 1650 (4GB DDR5)
Memory	8GB DDR4
Storage Type	Solid State Drive

Table 2. Software Components

COMPONENT	SPECIFICATION
Operating System	Windows 11
Programming Language	Python 3.9.2
Deep Learning Framework	PyTorch 1.13.1
Framework	Detectron 2

Our hardware and software configuration were carefully selected to ensure optimal performance and efficiency throughout the research process. The combination of a powerful GPU, a well-equipped CPU, and a rich software ecosystem enabled us to train and evaluate the Mask R-CNN model effectively, contributing to the success of our study.

5.2 OUTPUT IMAGES

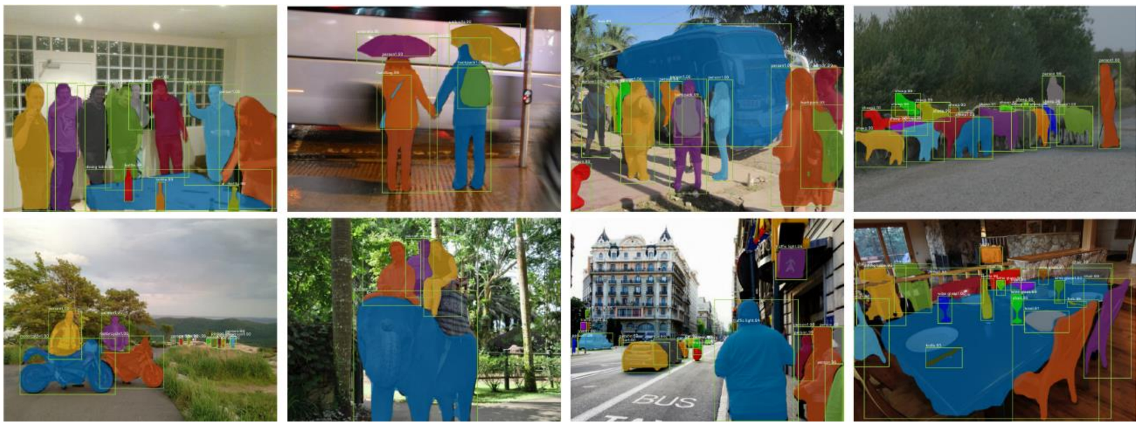

Figure 4. Mask R-CNN *results on the* ***COCO test set.*** *These results are based on* ***ResNet-101,*** *achieving a mask* ***AP of 35.7*** *and running at 5 fps. Masks are shown in color, and bounding box, category, and confidences are also shown.*

Figure 5. *More results of* ***Mask R-CNN*** *on* ***COCO test images,*** *using* ***ResNet-101-FPN*** *and running at 5 fps, with 35.7 mask AP*

Figure 6. *Keypoint detection results on **COCO** test using **Mask R-CNN (ResNet-50-FPN),** with person segmentation masks predicted from the same model. This model has a keypoint **AP of 63.1 and runs at 5 fps.***

5.3 Limitations and Drawbacks

Mask R-CNN is computationally intensive, which may hinder real-time applications. The processing of multiple regions of interest (RoIs) with complex masks and features can strain hardware resources.

Generating masks for each RoI requires significant memory. This might be challenging for systems with limited memory capacity.

Implementing and fine-tuning Mask R-CNN can be challenging, especially for users with limited experience in deep learning and computer vision.

Like other deep learning approaches, Mask R-CNN's performance heavily depends on the availability of high-quality, labelled training data, which can be resource-intensive to collect and annotate.

The performance of Mask R-CNN can be sensitive to hyperparameters and training settings, making it crucial to experiment and fine-tune these settings.

Mask R-CNN might struggle with instances that are noisy or overlap significantly, potentially leading to errors in segmentation.

Despite improvements, real-time applications may still be challenging due to the computational demands of Mask R-CNN.

The mask computation on the top 100 detection boxes introduces a small overhead compared to the Faster R-CNN counterpart, which could impact latency in certain applications.

While Mask R-CNN has demonstrated state-of-the-art results in instance segmentation, addressing these limitations is essential for broader and more practical use cases.

6. CONCLUSION AND FUTURE ENHANCEMENTS

Mask R-CNN stands as a powerful framework for instance segmentation and object detection, offering flexibility, compatibility, and robust performance. Developed by Facebook AI Research (FAIR), it extends the Faster R-CNN model with a mask prediction branch and incorporates the RoIAlign layer for precise mask prediction. Our extensive experiments and enhancements validate Mask R-CNN's effectiveness, demonstrating its superiority over other state-of-the-art models in instance segmentation. Its compatibility with various techniques makes it adaptable and extendable for detection and segmentation tasks. While Mask R-CNN's achievements are noteworthy, it presents challenges in terms of computational intensity, memory requirements, and sensitivity to hyperparameters.

Future enhancements offer promising directions for research and development:

Efficiency Optimization: Real-time applications, such as autonomous vehicles and robotics, require efficient model variants that balance accuracy and speed, optimizing inference times and resource utilization.

Data Augmentation Techniques: Mitigating limited labelled data challenges can be achieved through advanced data augmentation techniques, such as unsupervised data augmentation, generative adversarial networks, and domain adaptation, enhancing model generalization.

Noise Handling Strategies: Addressing noisy annotations or ambiguous instances is vital to ensure segmentation accuracy. Implementing noise handling strategies, such as filtering out annotation noise or uncertainty, will enhance the model's reliability.

Transfer Learning: Leveraging pre-trained models and knowledge transfer techniques can significantly improve Mask R-CNN's performance, especially in domains with limited annotated data.

Multi-Object Tracking: Enabling multi-object tracking capabilities in Mask R-CNN is essential for applications like surveillance, human-computer interaction, and autonomous navigation. Achieving this through extensions or integration with state-of-the-art tracking algorithms will expand its utility to a broader range of applications.

Mask R-CNN's robustness and adaptability, along with these proposed enhancements, position it as a valuable tool in the field of computer vision. By addressing its limitations and advancing its capabilities, we can continue to unlock its full potential for real-world challenges and applications.

7. REFERENCES

1. Hongjie He, Hongzhang Xu, Ying Zhang, Kyle Gao, Huxiong Li, Lingfei Ma, Jonathan Li, Mask R-CNN based automated identification and extraction of oil well sites, International Journal of Applied Earth Observation and Geoinformation, Volume 112, 2022, 102875, ISSN 1569-8432, https://doi.org/10.1016/j.jag.2022.102875.

2. Ayesha Munira Chowdhury, Sungwoo Moon, Generating integrated bill of materials using mask R-CNN artificial intelligence model, Automation in Construction, Volume 145, 2023, 104644, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2022.104644.

3. Donghyeon Kim, ChiUng Ko, Donggeun Kim, Method for detecting tree ring boundary in conifers and broadleaf trees using Mask R-CNN and linear interpolation, Dendrochronologia, Volume 79, 2023, 126088, ISSN 1125-7865, https://doi.org/10.1016/j.dendro.2023.126088.

4. Wenhao Lai, Feng Hu, Xixi Kong, Pengcheng Yan, Kai Bian, Xiangxiang Dai, The study of coal gangue segmentation for location and shape predicts based on multispectral and improved Mask R-CNN, Powder Technology, Volume 407, 2022, 117655, ISSN 0032-5910, https://doi.org/10.1016/j.powtec.2022.117655.

5. Deqiang He, Rui Ma, Zhenzhen Jin, Ruochen Ren, Suiqiu He, Zaiyu Xiang, Yanjun Chen, Weibin Xiang, Welding quality detection of metro train body based on ABC mask R-CNN, Measurement, Volume 216, 2023, 112969, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2023.112969.

6. R. Muztaba, H.L. Malasan, M. Djamal, Deep learning for crescent detection and recognition: Implementation of Mask R-CNN to the observational Lunar dataset collected with the Robotic Lunar Telescope System, Astronomy and Computing, Volume 45, 2023, 100757, ISSN 2213-1337, https://doi.org/10.1016/j.ascom.2023.100757.

7. Xiaofeng Qu, Jiajun Wang, Xiaoling Wang, Yike Hu, Tuocheng Zeng, Tianwen Tan, Gravelly soil uniformity identification based on the optimized Mask R-CNN model, Expert Systems with Applications, Volume 212, 2023, 118837, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.118837.

8. Lin Du, Xuemin Lu, Huazhou Li, Automatic fracture detection from the images of electrical image logs using Mask R-CNN, Fuel, Volume 351, 2023, 128992, ISSN 0016-2361,

https://doi.org/10.1016/j.fuel.2023.128992.

9. Yangyang Cao, Zuoxi Zhao, Yuan Huang, Xu Lin, Shuyuan Luo, Borui Xiang, Houcheng Yang, Case instance segmentation of small farmland based on Mask R-CNN of feature pyramid network with double attention mechanism in high resolution satellite images, Computers and Electronics in Agriculture, Volume 212, 2023, 108073, ISSN 0168-1699,

https://doi.org/10.1016/j.compag.2023.108073.

10. Ye Zhang, Yunlin Ma, Yanlong Li, Lifeng Wen, Intelligent analysis method of dam material gradation for asphalt-core rock-fill dam based on enhanced Cascade Mask R-CNN and GCNet, Advanced Engineering Informatics, Volume 56, 2023, 102001, ISSN 1474-0346, https://doi.org/10.1016/j.aei.2023.102001.

8. BASE ARTICLE

https://paperswithcode.com/paper/mask-r-cnn

https://github.com/facebookresearch/detectron2