

Final Project
Object Detection Mask R-CNN
Kelas MK Visi Komputer CII4Q3-IF-43-PIL-IS02



1301194008
Muhammad Sulthon Asramanggala

Program Studi Sarjana Informatika
Fakultas Informatika
Universitas Telkom
Bandung
2023

1. Review Paper

1.1. Judul

Mask R-CNN

1.2. Penulis

Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick

1.3. Abstrak

"Mask R-CNN" adalah algoritma pengklasifikasi objek yang digunakan dalam pengolahan citra dan computer vision. Algoritma ini dikembangkan oleh Kaiming He, Georgia Gkioxari, Piotr Dollár, dan Ross Girshick pada tahun 2017 dan diterbitkan dalam jurnal "IEEE Transactions on Pattern Analysis and Machine Intelligence". Mask R-CNN menggabungkan teknik pengklasifikasi objek dengan teknik deteksi objek dan juga menambahkan kapabilitas untuk melakukan segmentasi objek dalam citra. Ini dilakukan dengan menambahkan lapisan "mask" pada arsitektur Faster R-CNN yang sudah ada. Keuntungan dari Mask R-CNN adalah kemampuannya untuk secara akurat mengidentifikasi dan menandai objek yang terdapat dalam citra, yang dapat digunakan dalam aplikasi seperti pemetaan lingkungan, pengenalan objek, dan editing citra.

1.4. Pendahuluan

Dalam paper "Mask R-CNN", penulis mengidentifikasi masalah utama dalam deteksi objek dan segmentasi objek yaitu ketidakmampuan untuk secara akurat mengidentifikasi dan menandai objek yang terdapat dalam citra. Mereka menyatakan bahwa algoritma yang ada saat ini sering kali mengalami kesulitan dalam mengatasi masalah ini, terutama ketika objek tersebut beroverlap atau terpotong. Untuk mengatasi masalah ini, penulis mengembangkan Mask R-CNN, yang menggabungkan teknik pengklasifikasi objek dengan teknik deteksi objek dan juga menambahkan kapabilitas untuk melakukan segmentasi objek dalam citra. Ini dilakukan dengan menambahkan lapisan "mask" pada arsitektur Faster R-CNN yang sudah ada. Hasil dari paper ini menunjukkan bahwa Mask R-CNN mampu meningkatkan akurasi dalam mengidentifikasi dan menandai objek yang terdapat dalam citra dibandingkan dengan algoritma lain yang ada saat ini. Penulis menunjukkan hasil yang baik pada benchmark COCO dataset, yang menunjukkan bahwa Mask R-CNN mampu meningkatkan akurasi pengklasifikasi objek dan segmentasi objek dibandingkan dengan algoritma lain yang ada saat ini.

1.5. Related Work

Dalam paper "Mask R-CNN" penulis menyimpulkan beberapa algoritma terdahulu yang digunakan untuk deteksi objek dan segmentasi objek, diantaranya :

- a. R-CNN (Region-based CNN) yang menggunakan CNN untuk melakukan pengklasifikasi objek setelah melakukan deteksi objek menggunakan metode sliding window.
- b. Fast R-CNN yang menggunakan metode yang sama dengan R-CNN namun meningkatkan efisiensi dengan menggunakan teknik RoI pooling.
- c. Faster R-CNN yang menggabungkan teknik deteksi objek dengan pengklasifikasi objek menjadi satu algoritma dengan menambahkan lapisan Region Proposal Network (RPN)
- d. YOLO (You Only Look Once) yang menggunakan teknik deteksi objek yang berbeda dengan R-CNN dan Faster R-CNN dengan menggunakan satu jalur konvolusi untuk melakukan deteksi objek dan pengklasifikasi objek sekaligus.

Dengan kata lain Mask R-CNN merupakan evolusi dari Faster R-CNN dengan menambahkan kapabilitas segmentasi objek. Dalam paper "Mask R-CNN" juga

disebutkan menyebutkan beberapa metode segmentasi objek yang ada sebelumnya seperti FCN (Fully Convolutional Networks) dan DeepLab, namun menyatakan bahwa Mask R-CNN menawarkan akurasi yang lebih tinggi dan efisiensi yang lebih baik dibandingkan dengan metode tersebut.

1.6. Metode

Metode dalam paper “Mask R-CNN” dijelaskan cara kerja algoritma Mask R-CNN dalam menyelesaikan masalah deteksi objek dan segmentasi objek. Mask R-CNN terdiri dari 3 lapisan utama, diantaranya :


- a. Lapisan Region Proposal Network (RPN) yang digunakan untuk mendeteksi daerah potensial dari objek yang terdapat dalam citra.
- b. Lapisan pengklasifikasi objek yang digunakan untuk mengklasifikasikan objek yang telah dideteksi oleh RPN.
- c. Lapisan mask yang digunakan untuk melakukan segmentasi objek dengan menentukan batas objek secara pixel-level.



Proses deteksi objek dilakukan dengan mengambil citra sebagai input dan menggunakan RPN untuk menentukan daerah potensial dari objek yang terdapat dalam citra. Kemudian, lapisan pengklasifikasi objek digunakan untuk mengklasifikasikan objek yang telah dideteksi oleh RPN. Setelah itu, lapisan mask digunakan untuk melakukan segmentasi objek dengan menentukan batas objek secara pixel-level.


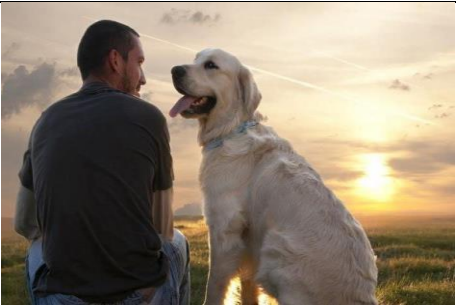


Teknik yang digunakan dalam algoritma Mask R-CNN seperti RoI Align yang digunakan untuk meningkatkan akurasi dalam proses pengklasifikasi objek dan teknik pengenalan objek yang diterapkan pada lapisan mask. Secara keseluruhan, Mask R-CNN menawarkan solusi untuk masalah deteksi objek dan segmentasi objek dengan menggabungkan teknik pengklasifikasi objek dengan teknik deteksi objek dan juga menambahkan kapabilitas untuk melakukan segmentasi objek dalam citra.

1.7. Eksperimen






Dalam eksperimen yang dilakukan pada final project kali ini penulis melakukan melakukan eksplorasi pada salah satu source code yang telah tersedia dalam situs “<https://paperswithcode.com>” yang berkaitan dengan paper “Mask R-CNN” dengan mengganti dataset yang telah ada dengan 10 gambar yang penulis cari dalam mesin pencari. Berikut merupakan gambar dataset inputan yang digunakan penulis :

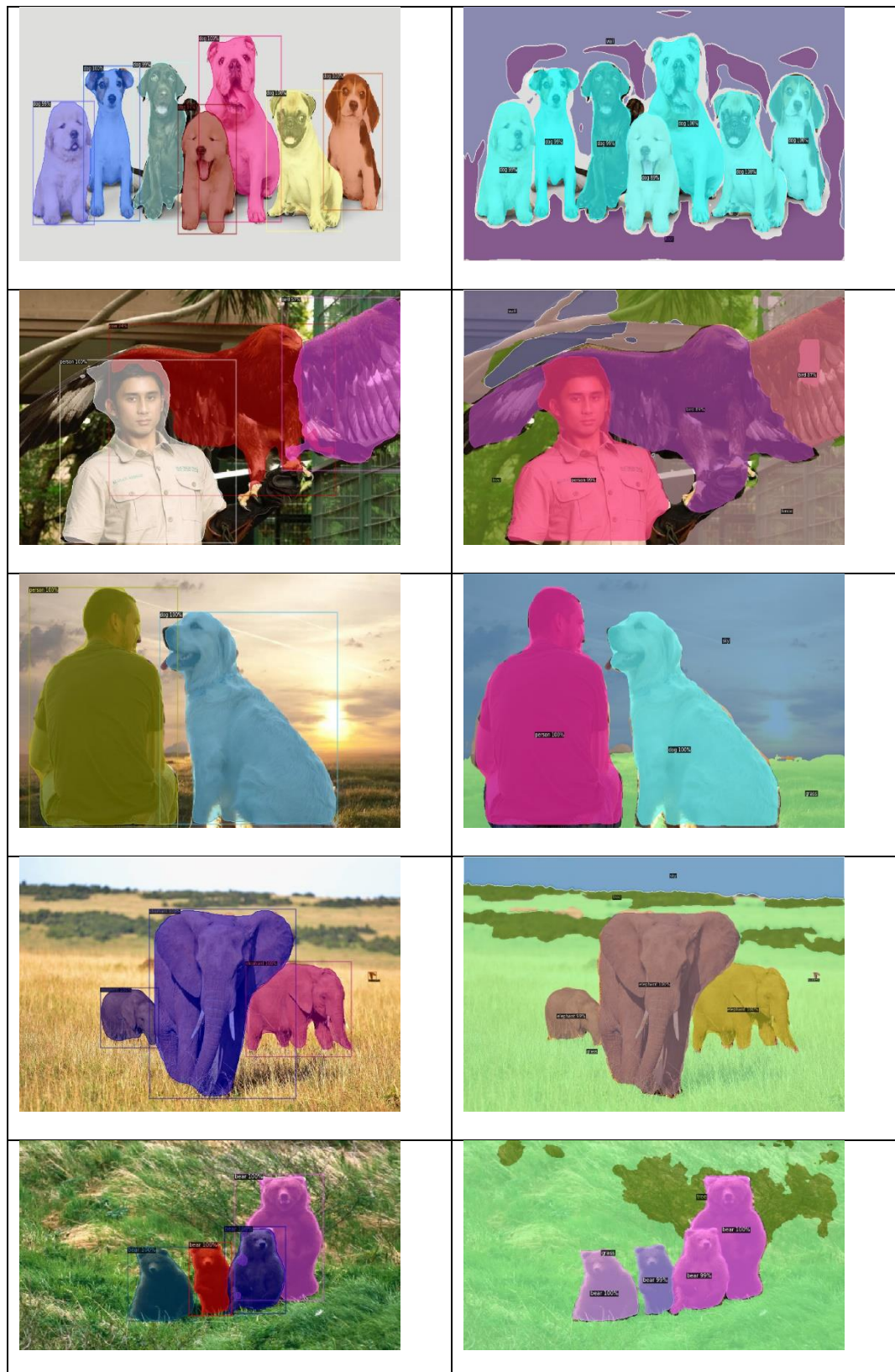
Input	Link
	Link dataset : dataset

	Link dataset : dataset
	
	
	
	

	Link dataset : dataset
	
	
	

Setelah dilakukan penginputan gambar maka gambar yang telah di input akan dilakukan proses prediksi gambar menggunakan model yang sudah tersedia pada benchmark COCO. Berikut merupakan output hasil object detection yang dilakukan :

Visualizer Prediction	Segmentation Model
	
	
	
	
	



Selain dengan gambar penulis juga melakukan eksperimen dengan menggunakan inputan video dari youtube. Hasil yang output yang diharapkan adalah cuplikan video selama 6 detik disertai dengan penoptic segmentation dan berhasil mengidentifikasi car dan road pada 6 detik pertama


```
[81] #This is the video we're going to process
from IPython.display import YouTubeVideo, display
video = YouTubeVideo("jYc766bDCDo", width=500)
display(video)
```



1.8. Kesimpulan

Dengan menambahkan kemampuan untuk melakukan instance segmentation, yaitu menghasilkan segmen mask untuk setiap objek yang terdeteksi, pada arsitektur deteksi objek "Faster R-CNN" yang populer, dapat meningkatkan akurasi deteksi dan segmentasi objek secara signifikan. Model "Mask R-CNN," menunjukkan hasil yang baik dalam berbagai dataset uji coba dan aplikasi. Bisa dilihat dari perbandingan dalam table output gambar yang dideteksi bahwa dengan menambahkan segmentation semua objek yang terdapat dalam gambar dalam di deteksi.

1.9. Link Penting

Source code :

<https://drive.google.com/file/d/1r3lEB1x9ZtpjdBwuBryGmRPTBjqm4HmD/view?usp=sharing>

Dataset :

<https://drive.google.com/drive/folders/1Yl-xkJaz-cmHGRVRMiwBhUydL7uVYscJ?usp=sharing>

Ouput :

https://drive.google.com/drive/folders/153KX97jE_MsbzfD1xBz1M_xXTVI0yS_M?usp=sharing

Paper :

<https://paperswithcode.com/method/mask-r-cnn>