

*Examine the effectiveness of the mask R-CNN method, and
keypoint R-CNN method in image object detection*

**Menguji keefektifan metode mask R-CNN, dan metode keypoint
R-CNN dalam deteksi objek citra**

Faris Eka Darmawan¹, Siti Khoerotul Magfiroh², Mutiara Khansa Hanifah³, Muhammad Saman⁴

¹ Program Studi Informatika, Universitas Muhammadiyah Semarang, Semarang, Indonesia

Info Artikel

Riwayat Artikel:

Diterima 9 Maret 2024

Perbaikan 15 Juni 2024

Disetujui 30 Juli 2024

Keywords:

Plant Disease
Classification
VGG16
ResNet34
CNN

ABSTRAK

Objek Detection merupakan tantangan dalam bidang computer vision yang mendukung aplikasi seperti pengenalan objek, pengenalan pola, dan analisis citra medis. Dalam penelitian ini, kami membandingkan kinerja dua metode utama dalam deteksi objek: Mask R-CNN, dan keypoint R-CNN. Metode R-CNN didasarkan pada pembuatan proposal wilayah menggunakan jaringan proposal wilayah (RPN), yang kemudian diproses oleh jaringan konvolusional untuk klasifikasi dan regresi. Oleh karena itu, Mask R-CNN mengintegrasikan kemampuan segmentasi instan dengan memberikan pemahaman yang lebih mendalam tentang objek. Di sisi lain, Keypoint R-CNN menambahkan dimensi pada deteksi objek dengan menentukan titik kunci atau landmark pada objek. Dimana dalam objek detection menggunakan sekitar 5000 gambar untuk implementasi metode Mask R-CNN dan Keypoint R-CNN. Evaluasi didasarkan pada akurasi deteksi, kecepatan eksekusi, dan kebutuhan sumber daya komputasi.

ABSTRACT

Object Detection is a challenge in the field of computer vision that supports applications such as object recognition, pattern recognition, and medical image analysis. In this study, we compare the performance of two main methods in object detection: Mask R-CNN, and keypoint R-CNN. The R-CNN method is based on generating region proposals using region proposal networks (RPN), which are then processed by convolutional networks for classification and regression. Therefore, Mask R-CNN integrates instant segmentation capabilities by providing a deeper understanding of the object. On the other hand, Keypoint R-CNN adds dimension to object detection by determining key points or landmarks on the object. Where in object detection using about 5000 images for the implementation of Mask R-CNN and Keypoint R-CNN methods. The evaluation is based on detection accuracy, execution speed, and computational resource requirements.

Ini adalah artikel akses terbuka di bawah lisensi CC BY-SA.



Penulis Korespondensi:

Faris Eka Darmawan
Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Semarang
Jl. Kedungmundu No. 18, Semarang, Jawa Tengah, Indonesia
Email: farisdermawan63@gmail.com

1. PENDAHULUAN

Teknologi deteksi objek pada pengolahan citra digital serta *computer vision* sudah berkembang sangat pesat serta memiliki peran yang penting dalam kecerdasan buatan atau *Artifical Intelligence*, deteksi objek bahkan sudah banyak diterapkan untuk berbagai kepentingan. Deteksi objek merupakan teknologi yang sedang dikembangkan, dan banyak peneliti tertarik pada pengembangan lebih lanjut untuk memaksimalkan kinerja pendeteksian objek. Pengenalan objek juga merupakan teknologi komputer yang berkaitan dengan citra digital dan pengolahan citra yang digunakan untuk mengenali objek seperti orang, bangunan, dan mobil. Pendeteksian objek akan dilakukan berdasarkan pada dataset yang diberikan dengan beberapa contoh barang yang terdapat di dalamnya seperti orang, meja, kursi, pisang, mobil dan motor. Hal tersebut berdasarkan analogi yang akan digunakan pada penelitian ini.

Computer Vision merupakan teknologi yang menyerupai kerja penglihatan manusia. Sehingga suatu mesin komputer bisa mengenali dan memahami suatu gambar. Salah satu aspek yang ada pada *Computer Vision* adalah *Object Detection*. Dengan adanya *Computer Vision*, teknologi ini dapat digunakan sebagai pendeteksian objek berdasarkan analogi yang akan digunakan. Data pelatihan dan data pengujian didapatkan melalui coco dataset sehingga memudahkann untuk import datasetnya. Berdasarkan hal di atas, pada penelitian ini akan menguji sebuah gambar yang didalamnya terdapat analogi sudah dibuat dan akan dilakukan deteksi objek berdasarkan analogi tersebut melalui google collab. Penelitian ini menggunakan algoritma RCNN dengan metode *Mask R-CNN*.

Ada banyak metode untuk melakukan *object detection* yang sudah banyak dikembangkan salah satunya adalah *Region Convolutional Neural Network*. Dimana dalam RCNN terdapat beberapa model contohnya yaitu *Mask R-CNN* dan *Keypoint R-CNN*. *Keypoint R-CNN* adalah model deteksi objek yang berfokus pada estimasi titik kunci (*keypoint*) pada objek dalam gambar atau citra, sedangkan *Mask R-CNN* merupakan *Deep Neural Network* yang mempunyai tujuan deteksi objek pada bidang *instance segmentation* di *machine learning* atau *computer vision*. *Mask R-CNN* memungkinkan untuk mengidentifikasi dan memisahkan objek yang berbeda dalam gambar dengan berdasarkan *instance segmentation*. *Keypoint R-CNN* dan *Mask R-CNN* masih ada hubungan dalam deteksi objek Dimana keduanya merupakan turunan dari *Region Convolutional Neural Network* dan juga menggunakan *resnet* untuk meningkatkan kecepatan serta akurasi.

Beberapa penelitian terkait yaitu rancangan yang dibuat dapat membantu pihak supermarket untuk melakukan monitoring pelanggan dalam melakukan pembayaran dengan menggunakan metode *Mask R-CNN*. Dan mendapatkan akurasi sebesar 95% yang diuji menggunakan video. Selanjutnya penelitian yang dibuat bertujuan untuk mendeteksi dan serta menghitung objek kendaraan yang melewati suatu lalu lintas pada jalan sesuai dengan pembagian terstruktur mengenai jenis-jenis kendaraan. Dan mendapatkan akurasi sebesar 90.8% rata-rata objek yang dideteksi. Kemudian penelitian yang dibuat digunakan sebagai keamanan dimana plat nomor kendaraan seorang kriminal dapat dideteksi secara otomatis melalui sistem yang dapat memberitahu pihak berwenang secara otomatis. Dan pada perancangan ini mendapatkan akurasi sebesar 73.8% dan 74.2% untuk keseluruhan karakter pada plat nomor kendaraan.

2. LANDASAN TEORI

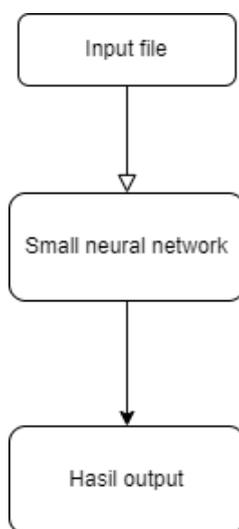
2.1 Object Detection

Algoritma R-CNN atau *Regional-based Convolutional Neural Network* adalah bagian dari CNN (*Convolutional Neural Network*) yang mengacu pada letak wilayah citra yang akan dikelompokkan. Algoritma R-CNN dikembangkan oleh Ross Girshick dan efektif dalam menemukan lokasi objek, mendeteksi objek, dan segmentasi objek. R-CNN membagi gambar menjadi beberapa region dan kemudian menghitung fitur untuk setiap region tersebut.

Kemudian, sebuah classifier digunakan untuk mengklasifikasikan apakah region tersebut mengandung objek atau tidak. R-CNN memiliki beberapa varian seperti Fast R-CNN, Faster R-CNN, Mask R-CNN, dan Keypoint R-CNN. Fast R-CNN menggabungkan *proses region proposal* dan *feature extraction* menjadi satu proses. Faster R-CNN memperkenalkan *Region Proposal Network (RPN)* yang memungkinkan proses region proposal menjadi lebih cepat. Mask R-CNN menggabungkan Faster R-CNN dengan sebuah layer baru untuk memprediksi segmentation mask dari tiap region. Sedangkan Keypoint R-CNN menggabungkan Mask R-CNN dengan sebuah layer baru untuk memprediksi *keypoint* dari objek yang terdeteksi.

2.2 Region Convolutional Neural Network

R-CNN yang merupakan turunan dari *Convolution Neural Network* yang dimana bisa digunakan pada data gambar. CNN dapat digunakan untuk mendeteksi suatu objek yang ada pada sebuah image atau gambar. CNN merupakan sebuah teknologi yang terinspirasi dari cara manusia menghasilkan penglihatan, yaitu kemampuan mata manusia untuk melihat, yang kemudian diproses oleh otak manusia hingga menghasilkan gambar yang terlihat.



Gambar 1. Proses pada metode R-CNN

Setiap gambar yang akan diproses frame demi frame akan menjadi kecil dan akan menghasilkan gambar terpotong dengan konvolusi yang sama. Setiap gambar yang diperoleh dari konvolusi kemudian digunakan sebagai masukan untuk membuat representasi fitur. Hal ini memberikan CNN kemampuan untuk mengenali suatu objek, di mana pun objek tersebut muncul.

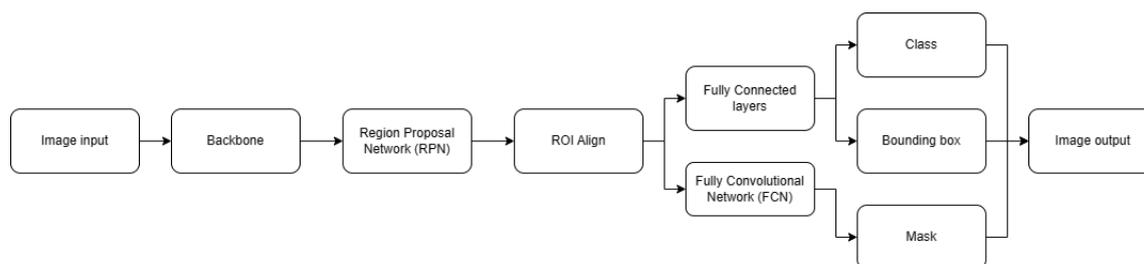
Proses ini dilakukan pada seluruh bagian dari setiap gambar kecilnya, dengan memanfaatkan filter yang sama. Dengan kata lain, setiap bagian gambar akan memiliki weight sharing yang sama, dan jika ada bagian yang terlihat menarik maka akan ditandai dengan bagian itu sebagai *object of interest*. R-CNN muncul karena banyaknya potongan gambar yang harus diproses oleh CNN. Regions based CNN terdiri dari 3 tahap:

1. Mencari region atau bagian gambar yang mungkin sebuah objek.
2. Tiap region kemudian dijadikan input untuk CNN sebagai ekstraktor dari tiap region.
3. Setiap fitur yang dihasilkan, kemudian menjadi input untuk SVM.

2.3 Mask Region-Convolutional Neural Network

Mask RCNN adalah sebuah framework untuk melakukan segmentasi objek yang merupakan perkembangan dari metode sebelumnya, yaitu Faster R-CNN. Metode ini menambahkan fitur yang mampu melakukan prediksi terhadap dimana kemungkinan sebuah objek berada dan memberikan mask atau lapisan pada objek. Pada Mask RCNN, diciptakan sebuah layer baru untuk memprediksi segmentation mask dari tiap *Region of Interest (ROI)* yang berjalan secara paralel dengan layer untuk klasifikasi dan *bounding box regression* yang sudah ada sebelumnya.

Berikut adalah arsitektur Mask RCNN:



Gambar 2. Arsitektur dari Mask RCNN

Pada di atas menunjukkan gambar arsitektur Mask R CNN memiliki langkah-langkah seperti berikut:

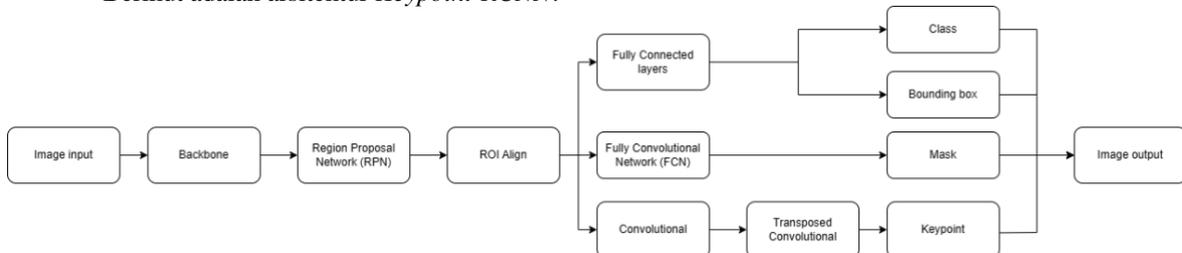
1. Proses dimulai dengan tahapan *Image Input*, yaitu memasukkan gambar yang akan dianalisis ke dalam model. Gambar dapat berupa beberapa objek dari berbagai kelas.
2. *Backbone*, Tahap ini memanfaatkan *pre-trained* CNN seperti ResNet atau VGG16 untuk mengekstrak fitur tingkat tinggi dari gambar input. Ini bertindak sebagai ekstraktor fitur, mengidentifikasi tepi, tekstur, dan pola yang relevan dengan pengenalan objek.
3. Region Proposal Network (RPN), yaitu pencarian yang secara cepat lokasi objek pada suatu gambar. Lalu mengusulkan wilayah persegi panjang (bounding box) yang kemungkinan besar berisi objek tadi, cara kerjanya yaitu menganalisis patch kecil peta fitur dan memprediksi kemungkinan setiap patch berisi objek dan kelas spesifiknya (misalnya, orang, mobil, kucing).
4. Region of Interest (ROI) Align, Setiap bounding box yang diusulkan (Region of Interest, ROI) dari RPN diperkecil dan diwarping ke ukuran tetap, memastikan input yang konsisten untuk tahap-tahap selanjutnya. Mudah-mudahan model menggunakan bounding box yang diusulkan untuk menghasilkan bounding box yang lebih akurat.
5. Beberapa implementasi Mask R-CNN menggunakan *Fully Connected Layers* setelah ROI Align untuk lebih menyempurnakan fitur yang diekstraksi khusus untuk setiap ROI. Tahap ini dapat meningkatkan kinerja dalam kasus tertentu.
6. Fully Convolutional Network (FCN), Tahap ini mengambil ROI yang diperkecil dan fitur terkaitnya dari langkah sebelumnya. Ini menggunakan jaringan saraf konvolusional untuk melakukan dua tugas independent:
 - a. Klasifikasi: Memprediksi probabilitas kelas untuk setiap ROI, menunjukkan jenis objek yang dikandungnya (misalnya, orang, mobil, anjing).
 - b. Prediksi Masker: Menghasilkan masker biner untuk setiap ROI. Setiap piksel dalam masker diberi probabilitas milik objek di ROI. Ini memungkinkan segmentasi halus bentuk objek.
7. *Class*, Berdasarkan cabang klasifikasi FCN, model menetapkan label kelas (misalnya, orang, mobil, anjing) ke setiap ROI dengan skor kepercayaan yang sesuai.
8. *Bounding Box*, Model menyempurnakan proposal bounding box awal dari RPN berdasarkan keluaran FCN. Hal ini menghasilkan bounding box yang lebih akurat yang mengelilingi objek yang terdeteksi dengan erat.
9. *Mask*, *binary mask* yang diprediksi oleh FCN untuk setiap ROI mendefinisikan bentuk dan batas yang tepat dari objek yang terdeteksi dalam bounding box. Ini memungkinkan segmentasi objek secara detail, membedakannya dari latar belakang dan objek lain dalam gambar.
10. *Image output*, Output akhir dari model Mask R-CNN adalah gambar asli dengan overlay bounding box untuk objek yang terdeteksi dan objek sebenarnya yang telah secara manual ditentukan, bersama dengan label *class* atau nama objek tersebut.

2.4 Keypoint Region-based Convolutional Neural Network

Keypoint-RCNN adalah sebuah model deteksi objek yang berbasis pembelajaran keypoint yang menghasilkan titik-titik penting (*keypoint*) dalam konteks object detection dan region proposal (RPN). Model ini menggunakan *Region Proposal Network (RPN)* untuk menghasilkan proposal wilayah dan menyederhanakan model dengan menggunakan keypoint head.

Keypoint-RCNN merupakan varian dari *R-CNN* yang mencoba menggabungkan kegunaan deteksi objek dan pengiraan keypoint dalam satu model. Model ini memiliki keunggulan tinggi dalam mengakurat dan memprediksi titik-titik penting dalam gambar, yang membuatnya berharga untuk berbagai aplikasi. *Keypoint-RCNN* dapat digunakan untuk mendeteksi titik-titik penting pada gambar manusia seperti wajah, bahu, pergelangan tangan, dan pergelangan kaki. Algoritma ini menggunakan loss function untuk mengoptimalkan deteksi titik penting pada gambar. *Keypoint-RCNN* merupakan pengembangan dari *Faster RCNN* dan *Mask RCNN*, dan dapat digunakan untuk berbagai aplikasi, seperti *Human Pose Estimation* dan pengendalian objek.

Berikut adalah arsitektur *Keypoint-RCNN*:



Gambar 3. Arsitektur dari Keypoint-RCNN

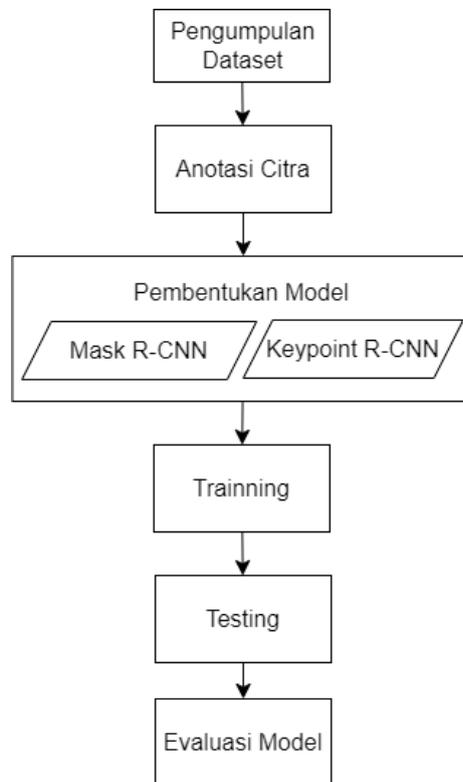
Pada di atas menunjukkan gambar arsitektur *Keypoint-RCNN* memiliki langkah-langkah seperti berikut:

1. Proses dimulai dengan tahapan *Image Input*, yaitu memasukkan gambar yang akan dianalisis ke dalam model. Gambar dapat berupa beberapa objek dari berbagai kelas.
2. *Backbone*, Tahap ini memanfaatkan *pre-trained CNN* seperti ResNet atau VGG16 untuk mengekstrak fitur tingkat tinggi dari gambar input. Ini bertindak sebagai ekstraktor fitur, mengidentifikasi tepi, tekstur, dan pola yang relevan dengan pengenalan objek.
3. *Region Proposal Network (RPN)*, yaitu pencarian yang secara cepat lokasi objek pada suatu gambar. Lalu mengusulkan wilayah persegi panjang (*bounding box*) yang kemungkinan besar berisi objek tadi, cara kerjanya yaitu menganalisis patch kecil peta fitur dan memprediksi kemungkinan setiap patch berisi objek dan kelas spesifiknya (misalnya, orang, mobil, kucing).
4. *Region of Interest (ROI) Align*, Setiap *bounding box* yang diusulkan (*Region of Interest, ROI*) dari RPN diperkecil dan diwarping ke ukuran tetap, memastikan input yang konsisten untuk tahap-tahap selanjutnya. Mudah-mudahan model menggunakan *bounding box* yang diusulkan untuk menghasilkan *bounding box* yang lebih akurat.
5. Beberapa implementasi Mask R-CNN menggunakan *Fully Connected Layers* setelah ROI Align untuk lebih menyempurnakan fitur yang diekstraksi khusus untuk setiap ROI. Tahap ini dapat meningkatkan kinerja dalam kasus tertentu.
6. *Fully Convolutional Network (FCN)*, Tahap ini mengambil ROI yang diperkecil dan fitur terkaitnya dari langkah sebelumnya. Ini menggunakan jaringan saraf konvolusional untuk melakukan dua tugas independent:
 - a. *Klasifikasi*: Memprediksi probabilitas kelas untuk setiap ROI, menunjukkan jenis objek yang dikandungnya (misalnya, orang, mobil, anjing).
 - b. *Prediksi Masker*: Menghasilkan masker biner untuk setiap ROI. Setiap piksel dalam masker diberi probabilitas milik objek di ROI. Ini memungkinkan segmentasi halus bentuk objek.
7. *Class*, Berdasarkan cabang klasifikasi FCN, model menetapkan label kelas (misalnya, orang, mobil, anjing) ke setiap ROI dengan skor kepercayaan yang sesuai.
8. *Bounding Box*, Model menyempurnakan proposal *bounding box* awal dari RPN berdasarkan keluaran FCN. Hal ini menghasilkan *bounding box* yang lebih akurat yang mengelilingi objek yang terdeteksi dengan erat.
9. *Mask, binary mask* yang diprediksi oleh FCN untuk setiap ROI mendefinisikan bentuk dan batas yang tepat dari objek yang terdeteksi dalam *bounding box*. Ini memungkinkan segmentasi objek secara detail, membedakannya dari latar belakang dan objek lain dalam gambar.
10. *Convolutional*, Lapisan convolutional tambahan ini memproses fitur yang disempurnakan dari ROI.

11. *Tranposed Convolutional*, Lapisan convolutional transpose ini meningkatkan resolusi fitur, mempersiapkannya untuk tahap prediksi *keypoint*.
12. *Keypoint*, Tahap akhir memprediksi lokasi *keypoint* pada objek yang terdeteksi di dalam bounding box. Setiap titik kunci memiliki skor kepercayaan yang menunjukkan keakuratan prediksi.
13. *Image output*, Output akhir dari model Keypoint R-CNN adalah gambar asli dengan overlay bounding box untuk objek yang terdeteksi dan objek sebenarnya yang telah secara manual ditentukan, bersama dengan label *class* atau nama objek tersebut seperti Gambar 6.

3. METODE PENELITIAN

Pada gambar di bawah ini dapat dilihat flowchart dalam deteksi objek menggunakan algoritma *mask rcnn* dan *keypoint rcnn*. Dalam gambar dibawah pada tahap awal akan menimport dataset ataupun mengunduhnya, setelah itu akan dilakukan anotasi citra berdasarkan gambar yang akan diolah. Dalam anotasi citra juga akan dibuat *bounding box*, proses tersebut selanjutnya ada prediksi model Dimana model apa saja yang akan digunakan. Langkah akhir barulah akan membuat evaluasi dalam perancangan program juga terdapat IoU untuk mengukur *bounding box* terhadap objek yang terdekatnya. Lebih lanjut flowchart di bawah Langkah pertama yaitu import library, kedua yaitu import dataset terdiri dari masukkan dataset, anotasi citra dan *pre-processing* data, kemudia ketiga ada prediksi dalam pemodelan dan langkah terakhir ada membuat model serta evaluasi untuk mengetahui akurasi pada model.



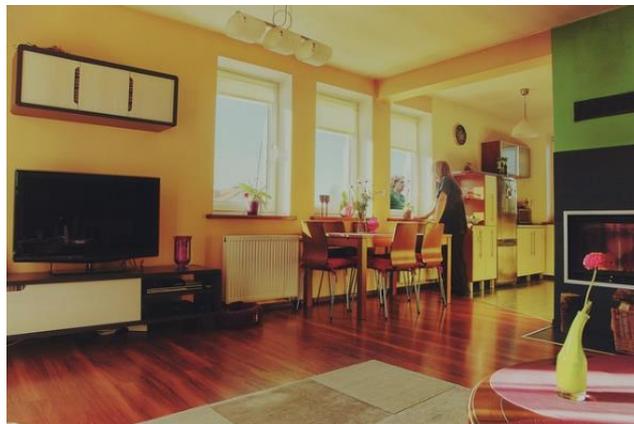
Gambar 4. Flowchat deteksi objek

3.1 Pengumpulan Dataset

Data yang digunakan pada penelitian ini menggunakan dataset yang tersedia dalam cocodataset tepatnya di http://images.cocodataset.org/annotations/annotations_trainval2017.zip. Dalam dataset tersebut berjumlah data latih yang digunakan pada penelitian ini sebanyak 5000 citra atau gambar. Di dalamnya terdapat berupa gambar *bus, truck, motorcycle, person, boat, chair, bear, traffic, bedroom* dan masih ada beberapa lagi. Dimana 5000 dataset tersebut akan digunakan untuk training sebelum dilakukan pemodelan.



Gambar 5. Contoh gambar bedroom



Gambar 6. Contoh gambar person, chair

Pada Gambar 5 dan 6 merupakan salah satu contoh gambar yang digunakan pada dataset, dalam penggunaan dataset pada penelitian ini secara langsung mengunduh melalui google collab tanpa perlu untuk unduh gambar sehingga bisa memudahkan untuk proses model.

3.2 Anotasi Citra

Setelah gambar selesai di proses maka selanjutnya adalah anotasi citra. Anotasi citra atau gambar adalah proses memberi label atau menandai gambar untuk memberikan informasi tambahan tentang objek yang ada didalamnya, biasanya dalam deteksi objek akan memberikan kotak pembatas atau *bounding box* beserta nama didalamnya. Proses anotasi citra atau pelabelan harus digunakan agar data latih yang digunakan memiliki nilai berdasarkan categories ataupun pemberian indeks pada gambar atau *image*. Contoh penggunaan anotasi citra yaitu pemberian indeks 1 untuk *chair* dan indeks 0 untuk *table*. Adapun penggunaannya menggunakan nama kategori misal ada gambar orang maka akan mencantumkan kategori *person* dalam label *categories*. Jadi dalam pelabelan dapat menggunakan angka jika gambar yang akan diseteksi hanya ada beberapa kategori, jika menggunakan selain angka juga bisa dengan gambar yang banyak. Dalam hal ini anotasi citra akan menunjukkan bounding box berwarna hijau untuk menunjukkan deteksi objek beserta namanya.

3.3 Pembentukan Model Metode Mask R-CNN dan Keypoint R-CNN

Pada penelitian ini metode Mask R-CNN dan Keypoint R-CNN di dalam google collab menggunakan device type T4 GPU untuk memperlancar pembacaan pada deteksi objek. Menggunakan torchvision untuk arsitektur model deteksi objek dan resnet 50 FPN. Torchvision merupakan salah satu library python di pytorch yang digunakan untuk *computer vision* dalam arsitektur model. Sedangkan model resnet 50 FPN merupakan model terbaru dari RCNN yang dikembangkan oleh pytorch agar lebih cepat dalam deteksi objek.

Pembentukan model *Mask R-CNN* dan *Keypoint R-CNN* terdapat layer yang masing-masing digunakan untuk pelatihan data nantinya. Dalam *Mask R-CNN* terdapat beberapa layer yang digunakan yaitu

backbone with FPN, Region Proposal Network (RPN), ROI Heads, dan Mask RCNN Predictor sedangkan *Keypoint R-CNN* menggunakan *Backbone with FPN, Feature Pyramid Network(FPN), Region Proposal Network(RPN), ROI Heads, dan Keypoint RCNN Predictor*.

3.4 Pelatihan Data

Pelatihan data dilakukan dengan mengekstraksi fitur dari setiap kelas untuk diklasifikasikan. Data pelatihan mencakup semua gambar di direktori. Jumlah iterasi yang dilakukan pada proses pelatihan sebanyak 5.000 iterasi. Saat pelatihan dijalankan, file bobot dihasilkan dalam format bobot. Weight Format adalah format file bobot saat melakukan proses pelatihan melalui *darknet*.

3.5 Rancangan Program

Pada saat perancangan program dilakukan pengujian bobot proses dengan menggunakan metode *Mask R-CNN* dan *Keypoint R-CNN* diperoleh hasil berupa deteksi kendaraan, barang, klasifikasi, dan komputasi. Dalam perancangan program terdapat program pendeteksi, program pendeteksi merupakan program yang dirancang untuk menguji bobot dalam mendeteksi citra menggunakan data yang akan diuji. Citra data diuji menggunakan dua metode yaitu *Mask R-CNN* dan *Keypoint R-CNN* dengan parameter referensi data latih berupa bobot hasil training. Keluaran dari program ini adalah hasil deteksi citra klasifikasi, serta prediksi akurasi dari data citra yang digunakan.

Proses selanjutnya yaitu perhitungan IoU, IoU adalah metrics yang digunakan untuk mengukur sejauh mana kotak pembatas atau *bounding boxes* tumpang tindih satu sama lain. Jika mendekati 1 maka luas area saling tumpang tindih diantara *bounding box* dan sebaliknya jika mendekati 0 maka dua *bounding box* hamper tidak atau bahkan tidak saling tumpang tindih.

4. HASIL DAN PEMBAHASAN

Uji coba dalam artikel ini menggunakan dataset yang terdapat 5000 citra atau gambar, pada dataset tersebut terdapat berbagai macam kategori. Kategori tersebut sebanyak 91 label *categories*. Di mana dalam dataset tersebut terdapat benda hidup seperti manusia, hewan dan tumbuhan. Dan juga ada benda mati contohnya mobil, bus, board, dan *traffic light*. Salah satu contoh hasil klasifikasi akan ditunjukkan pada gambar dengan masing-masing model yaitu *Mask R-CNN* dan *Keypoint R-CNN*. Pada bagian ini akan dilakukan proses menguji prediksi label, skor dan juga IoU pada data uji. Pengujian ini akan terdapat citra pada dataset dengan terdapat sekitar 91 label kategori.

4.1 Hasil dan Pembahasan *Mask R-CNN*

Mask RCNN adalah metode pertama yang akan digunakan untuk melakukan proses penelitian pada citra gambar. Dengan ukuran *resize* citra kurang lebih 600 x 400 pada hasil akhir.



Gambar 7. Hasil prediksi model metode mask R-CNN

Pada pembahasan tentang metode *Mask RCNN* akan dilakukan proses prediksi label, perhitungan skor dan juga berapa nilai IoU. Dalam Gambar 5 Hasil prediksi model metode *mask rcnn* terdapat beberapa label

dalam citra, seperti chair, person, dan table. Dalam hal ini metode *Mask Rcn* mampu menganalisa objek yang ada dalam citra, dengan prediksi label *chair*, *score* 0,997 dan juga untuk nilai IoU nya sebesar 0.000 dimana *bounding box* tidak terdapat tumpang tindih satu sama lain.

4.2 Hasil dan Pembahasan *Keypoint R-CNN*

Metode kedua yang digunakan adalah *Keypoint RCNN*, berdasarkan *Gambar 6 Hasil prediksi model metode keypoint rcnn* dengan menggunakan citra yang sama namun untuk hasil prediksi yang terdapat pada citra berbeda dengan metode *Mask RCNN*. Hasil prediksi dari metode ini adalah *person* dengan nilai *score* mencapai 0,95 dan nilai Io Unya adalah 0,81 di mana untuk *bounding box* saling tumpang tindih.



Gambar 8. Hasil prediksi model metode keypoint rcnn

Dalam metode *Keypoint RCNN* dapat mendeteksi emosi pada wajah, berdasarkan gambar diatas pada *bounding box* warna hijau mampu untuk menganalisa emosi pada wajah. Dimana dalam citra diatas memang mendeteksi *person* yang ada dalam citra.

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Pada penelitian ini didapatkan kesimpulan yaitu dalam menguji keefektifan *metode Mask R-CNN dan Keypoint R-CNN* lebih unggul *Mask R-CNN* dengan *score* 0,97 sedangkan *Keypoint R-CNN* memiliki *score* 0,95. Masing-masing prediksi memiliki nilai di atas 0,9 yang berarti metode *Mask R-CNN dan Keypoint R-CNN* cocok untuk digunakan dalam memprediksi citra atau gambar. Selanjutnya ada nilai IoU yang sebenarnya tidak berpengaruh terhadap hasil prediksi namun nilai IoU lebih ke apakah benda tersebut saling berdekatan atau tumpang tindih atau tidak. Metode *Mask RCNN dan Keypoint R--CNN* juga benar untuk *predicted label* pada gambar di atas walaupun pastinya ada kesalahan di dalamnya.

5.2. Saran

Dari hasil Penelitian ini, disarankan untuk dapat dilakukan penelitian lagi untuk melanjutkan atau mengembangkan penelitian ini seperti menambahkan jumlah data latih dan variasinya untuk mendapatkan hasil yang lebih baik dan akurat mengingat R-CNN bergantung pada jumlah data latih dan variasi data latih yang digunakan.

REFERENSI

- [1] Black, F. & Scholes, M., 1973. The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 81(3), PP. 637-659.
- [2] Learn OpenCV. (2023, December 10). Pemahaman sederhana tentang Mask RCNN. Learn OpenCV Pemahaman Sederhana tentang Mask RCNN (ichi.pro)
- [3] A. Primawati, I. Mutia, and D. Marlina, "Analisis Klasifikasi Populasi Ternak Kambing Dan Domba Dengan Model Convolutional Neural Network," *Fakt.Exacta*, vol. 14, no. 1, p. 22, 2021, doi:10.30998/faktorexacta.v14i1.8734.
- [4] Wicaksono, M. R. A., Setianingsih, C., & Saputra, R. E. (2023). Sistem Pendeteksi Sapi pada Peternakan dengan Citra Video UAV Menggunakan Metode Regions Based Convolution Neural Network. *e-Proceeding of Engineering*, 10(5), 4608. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/21197>
- [5] Gnanasekharan, V. dan Floros, J. D. (1995). Back propagation neural networks: Theory and applications for food science and technology. Dalam Charalambous, G. (ed.). *Food Flavors: Generation, Analysis and Process Influence*, hal 2151-2168. Elsevier Science B.V
- [6] Q. Lina, "Apa itu Convolutional Neural Network? | by QOLBIYATUL LINA | Medium.= <https://medium.com/@16611110/apa-itu-convolutional-neural-network836f70b193a4> (accessed Des. 14, 2023).
- [7] Ariya, C., & Lina, L. (2023). PERANCANGAN DETEKSI OBJEK PADA RAK TOKO MENGGUNAKAN METODE MASK RCNN. *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, 8(2), 295–299. <https://doi.org/10.51876/simtek.v8i2.213>
- [8] Learn OpenCV. (2023, December 10). Human Pose Estimation using Keypoint RCNN in PyTorch. Learn OpenCV. <https://learnopencv.com/human-pose-estimation-using-keypoint-rcnn-in-pytorch/>
- [9] Sompie, S., R.U.A., Wantania, B., B.M., & Kambey, F. D., "Penerapan Pendeteksian Manusia Dan Objek Dalam Keranjang Belanja Pada Antrian Di Kasir," *Jurnal Teknik Informatika*, vol. 15, no. 2, p. 101-108, 2020.
- [10] Mulyana, D. I., Aditya, A. C., Amelia, S., & Agustiansyah, S., "Implementasi Framework Mask R-CNN Object Detection API Dalam Mengklasifikasi Jenis Kendaraan Bermotor," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 4, no. 1, p. 63-72, 2022
- [11] Mesakh, A., "Sistem Pengenalan Plat Nomor Kendaraan Menggunakan Mask RCNN dan CNN," *Jurnal Ilmu Komputer dan Sistem Informasi*, 2022, vol. 10, no. 1