

Perancangan Sistem Deteksi Hewan Ternak Dengan Pendekatan Segmentasi Menggunakan MASK R-CNN

1st M Faiq Rofifi
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia

mfaiqr.student@telkomuniversity.ac.id

2nd Suryo Adhi Wibowo
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia

suryoadhiwibowo@telkomuniversity.ac.id

3rd Miftadi Sudjai3
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia

miftadi@telkomuniversity.ac.id

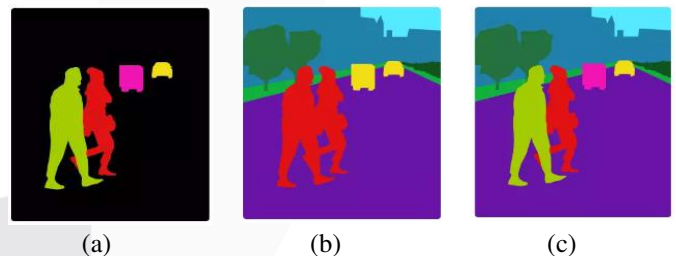
Abstrak — Industri peternakan menghadapi tantangan dalam pengawasan hewan ternak, khususnya dalam mendeteksi dan mengidentifikasi individu hewan secara akurat di tengah berbagai kondisi lingkungan yang kompleks. Salah satu tantangan utama adalah tumpang tindih objek atau oklusi, yang dapat menyebabkan kesalahan dalam deteksi, seperti ketika beberapa hewan terdeteksi dalam satu bounding-box. Dalam penelitian ini, teknologi *Computer Vision* digunakan untuk mengatasi masalah tersebut melalui pelatihan model *Mask R-CNN* yang diterapkan pada tiga kelas hewan ternak: sapi, domba, dan kuda. Pemilihan kelas hewan ini didasarkan pada prevalensi mereka dalam industri peternakan modern. Model dilatih dan diimplementasikan pada gambar yang mengandung hewan dari kelas yang sama maupun berbeda, serta diuji dalam kondisi oklusi untuk mengevaluasi kemampuan model dalam mendeteksi dan melakukan segmentasi objek di daerah yang sulit. Hasil pelatihan menunjukkan total *loss* sebesar 0,822, dengan akurasi deteksi objek yang mencapai 94,73% dan akurasi segmentasi sebesar 90,1%. Evaluasi lebih lanjut dilakukan untuk mengukur kualitas prediksi model, yang menghasilkan nilai *Mean Average Precision* (MAP) sebesar 0,427 untuk deteksi objek dan 0,364 untuk segmentasi. Tingkat kesalahan deteksi yang rendah juga tercatat, dengan sekitar 2 dari 5 objek terdeteksi secara salah.

Kata kunci— Oklusi, Mask R-CNN, *instance segmentation*, COCO Dataset

I. PENDAHULUAN

Pengawasan hewan ternak dalam industri peternakan merupakan tantangan yang kompleks, terutama dalam mendeteksi dan mengidentifikasi hewan di bawah kondisi lingkungan yang bervariasi, seperti oklusi dan interaksi antar hewan. Teknologi deteksi objek berbasis deep learning, seperti *instance segmentation* dengan Mask R-CNN, menawarkan solusi yang lebih canggih dibandingkan metode tradisional. Mask R-CNN, yang merupakan pengembangan dari R-CNN, telah menunjukkan peningkatan performa signifikan dengan *Mean Average Precision* (mAP) awal sebesar 36.4% [1]. Pengembangan lebih lanjut melalui PANet dengan hierarki bottom-up meningkatkan mAP menjadi 36.5%, dan penambahan network block serta

perhitungan IOU meningkatkan mAP hingga 39.1% [2]. Penelitian sebelumnya mendukung potensi besar teknologi ini. Model YOLOv5, misalnya, mampu mendeteksi sapi dengan *precision* mencapai 94.3% dan *recall* 92.5%, menunjukkan keandalannya dalam berbagai kondisi drone UAV [3]. Selain itu, YOLOv7 yang menggunakan kamera termal berhasil mendeteksi sapi dengan *presisi* 98% dan *recall* 67% [4], bahkan dalam kondisi gelap gulita. Model SheepInst juga menunjukkan hasil yang sangat baik untuk *Precision Livestock Farming* (PLF), dengan AP bounding-box sebesar 89.1%, AP mask 91.3%, dan AP batas 79.5% untuk deteksi dan segmentasi domba [5].



GAMBAR 1 (A).
(a) semantic segmentation (b) panoptic segmentation (c) Instance segmentation.

Berdasarkan hasil-hasil tersebut, dapat disimpulkan bahwa teknologi deep learning berbasis *instance segmentation* memiliki potensi besar untuk meningkatkan efisiensi dan akurasi dalam pemantauan hewan ternak. Namun, untuk mencapai hasil yang lebih optimal, penelitian lebih lanjut diperlukan untuk mengoptimalkan deteksi dan segmentasi objek. Oleh karena itu, penelitian ini bertujuan untuk merancang sistem deteksi hewan ternak berbasis Mask R-CNN yang diharapkan dapat memberikan hasil yang lebih akurat dan efisien dalam berbagai kondisi lingkungan peternakan.

II. KAJIAN TEORI

Menyajikan dan menjelaskan teori-teori yang berkaitan dengan variabel-variabel penelitian. Poin subjudul ditulis dalam abjad.

A. Dataset COCO(Common Objects in Context)

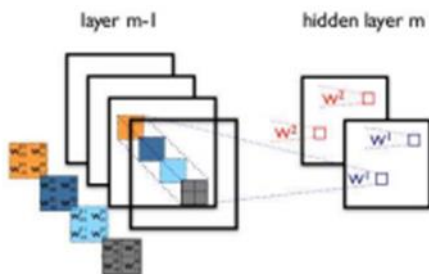
Dataset Microsoft *Common Objects in Context* (MS COCO) merupakan kumpulan data skala besar yang dirancang untuk meningkatkan teknologi pengenalan objek dengan mempertimbangkan pemahaman scene yang lebih luas. Dataset ini mencakup 328.000 gambar rumit dari kehidupan sehari-hari yang menampilkan 91 tipe objek umum dalam konteks alaminya. Objek-objek tersebut dilabeli dengan segmentasi per-instansi, menghasilkan total 2,5 juta instansi yang dilabeli untuk membantu dalam lokalisasi objek yang tepat [6].



GAMBAR 2 (A)
Contoh gambar segmentasi

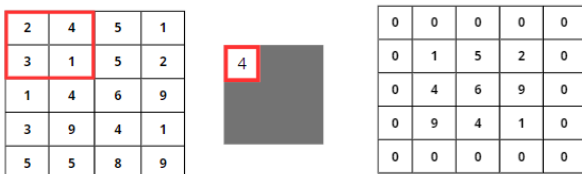
B. CNN (Convolutional Neural Network)

Convolutional Neural Network (CNN) adalah varian neural network yang khususnya efektif dalam memproses data gambar. Dibandingkan dengan model neural network konvensional, CNN memiliki struktur yang terdiri dari dua komponen utama: Feature Extraction Layer dan Fully-Connected Layer [7].



Gambar 1
(B) Konvolusi

Layer konvolusi dalam CNN adalah komponen kunci yang mendeteksi pola atau fitur khusus pada gambar dengan menerapkan filter kecil (kernel) yang melakukan operasi dot product pada setiap piksel, bergerak secara bertahap melintasi gambar [7].

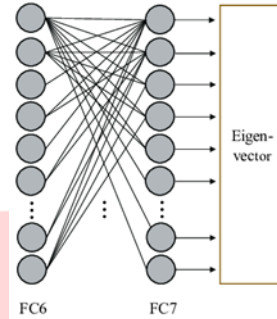


(a)

(b)

GAMBAR 2 (B).
Stride 1 piksel (a) Padding 1 (b)

1. Stride mengatur pergeseran filter saat konvolusi, mempengaruhi jumlah informasi yang diambil dan ukuran output gambar.
2. Padding menambahkan piksel di tepi gambar untuk memastikan filter menjangkau semua piksel dan mempertahankan dimensi output.

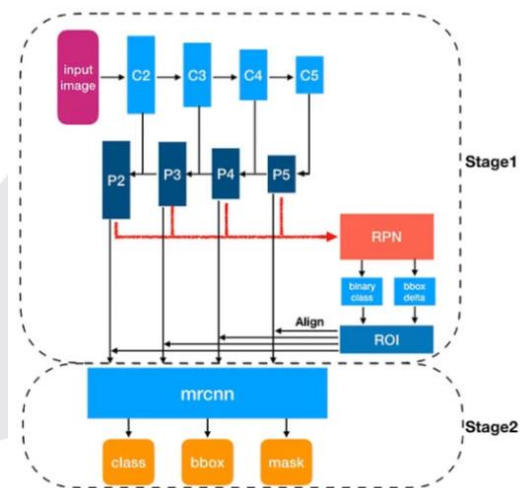


GAMBAR 3 (B)
Struktur FC6 dan FC7

Lapisan Fully Connected (FC), atau lapisan dense, adalah lapisan dalam jaringan saraf tiruan di mana setiap neuron terhubung penuh ke neuron di lapisan berikutnya, biasanya digunakan di akhir arsitektur jaringan untuk menghasilkan prediksi akhir.

C. Mask-RCNN

Mask R-CNN adalah pengembangan dari Faster R-CNN yang menambahkan segmentasi instance, dengan dua tahap utama: menghasilkan proposal region untuk objek dan memprediksi kelas, bounding-box, serta segmentasi piksel, didukung oleh backbone FPN [8].



GAMBAR 1 (C)
Arsitektur Mak-RCNN: FPN(P2,P3,P4,P5), RPN dan Prediksi

1. Backbone Mask R-CNN menggunakan Feature Pyramid Network (FPN) yang menggabungkan pendekatan bottom-up dan top-down untuk ekstraksi fitur, menghasilkan prediksi pada berbagai resolusi dengan integrasi skip dan lateral connections.
2. RPN dalam backbone FPN menentukan lokasi objek dengan menghasilkan proposal berdasarkan probabilitas dan perhitungan anchor melalui konvolusi 1x1, diterapkan pada berbagai resolusi.

- ROI Align mempertahankan informasi optimal dengan menggunakan interpolasi linear tanpa data quantization, membagi output menjadi grid untuk po

III. METODE

A. Proses Pengumpulan Data dan Integrasi ke Detectron 2

Dalam penelitian ini, data dari COCO dataset diambil dan diformat sesuai kebutuhan detectron 2 untuk pemrosesan model deteksi objek, yang menggunakan gambar dan file JSON beranotasi. Dataset difilter untuk hanya mencakup objek dengan tiga label: domba, sapi, dan kuda, guna memastikan relevansi dengan penelitian ini.



GAMBAR 1 (A)
Box diagram Pemrosesan Dataset

- Kami mengunduh dataset COCO, menyusun folder, mengekstraksi anotasi, dan menyaring label sapi, domba, dan kuda menggunakan cocomanager
- Skrip filter.py digunakan untuk menyaring data JSON berdasarkan input, output, dan kategori, kemudian dilanjutkan dengan mengunduh gambar uji sesuai data yang telah disaring.
- Setelah gambar diunduh, dataset didaftarkan ke Detectron menggunakan fungsi register coco instances dan metadata Catalog untuk menyiapkan data dan kelas untuk pelatihan dan evaluasi.

B. Proses Pelatihan Model

Sub bab ini menjelaskan proses pelatihan model menggunakan data yang telah difiltrasi, dikelompokkan, dan diregistrasi ke Detectron 2, dengan data pelatihan disimpan dalam folder "data training" dan mencakup beberapa konfigurasi dasar.

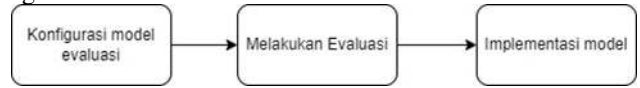
TABEL 1 (B).
Konfigurasi Pelatihan Menggunakan Detectron

Parameter	Nilai
cfg.DATASETS.TRAIN	("my dataset train1",)
cfg.DATALOADER.NUM WORKERS	2
cfg.MODEL.WEIGHTS	model zoo.get_checkpoint_url("COCOInstance Segmentation/ mask_rcnn R 101 FPN 3x .yaml")
cfg.SOLVER.IMS PER BATCH	2
cfg.SOLVER.BASE LR	0.00025
cfg.SOLVER.MAX ITER	300
cfg.MODEL.ROI HEADS.BATCH SIZE PER IMAGE	128
cfg.MODEL.ROI HEADS.NUM CLASSES	3

C. Proses Pelatihan Model

Proses inferensi menggunakan dataset evaluasi mencakup perhitungan AP, AR, dan analisis mAP untuk memastikan

performa model pada bounding box, prediksi kelas, dan segmentasi



GAMBAR 1 (C)
Box diagram Pemrosesan Dataset

- Model dilatih dan disimpan sebagai checkpoint, lalu diinferensi menggunakan dataset "eval" dengan hasil disimpan untuk analisis lebih lanjut.
- Evaluasi menghitung AP dan AR berdasarkan IoU, skala objek, dan jumlah objek, menghasilkan mAP untuk deteksi dan segmentasi yang dibandingkan dengan penelitian sebelumnya.
- Implementasi diuji pada video beragam kompleksitas dan occlusion, dengan evaluasi error prediksi untuk menilai kinerja model.

IV. HASIL DAN PEMBAHASAN

A. Hasil dan Analisis Pengambilan Dataset

Bab ini memaparkan hasil filtrasi kelas sapi, domba, dan kuda, termasuk anotasi JSON, dan unduhan gambar dari COCO.

```

[04/17 04:20:53 d2.data.build]: Removed 0 images with no usable annotations. 6161 images left.
[04/17 04:20:53 d2.data.build]: Distribution of instances among all 3 categories:
category | #instances | category | #instances | category | #instances
-----|-----|-----|-----|-----|-----
horse | 6567 | sheep | 9223 | cow | 8014
total | 23804
    
```

GAMBAR 1 (A)
Log dataset yang terdaftar pada Detectron 2

Log Gambar 1 (A) mencatat bahwa dataset pelatihan terdiri dari 6161 gambar, dengan 8014 objek kelas sapi, 9223 objek kelas domba, dan 6567 objek kelas kuda.

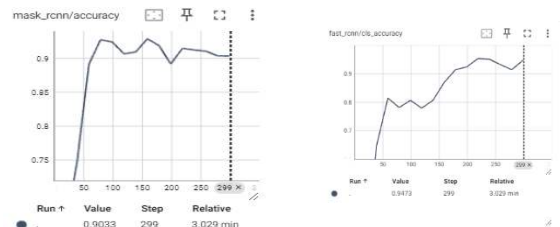
```

[04/17 04:26:57 d2.data.datasets.coco]: Loaded 267 images in COCO format from /content/dai
[04/17 04:26:57 d2.data.build]: Distribution of instances among all 3 categories:
category | #instances | category | #instances | category | #instances
-----|-----|-----|-----|-----|-----
horse | 272 | sheep | 354 | cow | 372
total | 998
    
```

GAMBAR 2 (A)
Log dataset yang terdaftar pada Detectron 2

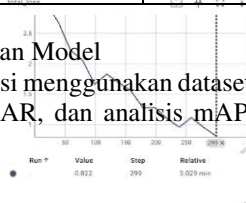
Log Gambar 3.2 mencatat bahwa dataset evaluasi terdiri dari 267 gambar, dengan 372 objek kelas sapi, 354 objek kelas domba, dan 272 objek kelas kuda.

B. Analisis Hasil Training Model



(a)

(b)



Run # Value Step Relative
0.9033 299 3.029 min

REFERENSI

(c)

GAMBAR 1 (B)

Mask RCNN akurasi (a) faster RCNN akurasi (b) Total Loss (c)

1. Grafik 1(B).b menunjukkan fluktuasi akibat minibatching, namun akurasi model meningkat stabil dengan momentum, mencapai 94,73% pada batch ke-299.
2. Grafik 1.(B).a menunjukkan akurasi segmentasi piksel meningkat signifikan dan stabil di atas 80%, mencapai 90,1% pada batch ke-79.
3. Grafik 1.(C).c menunjukkan total loss, termasuk loss kelas, box regresi, dan mask segmentasi, yang menurun stabil hingga mencapai 0,822.

C. Evaluasi Kinerja Model Deteksi Objek Menggunakan mAP

TABLE 1 (C)
Mean Average Precision (mAP)

Metrik	Nilai
mAP (Deteksi)	0.427
mAP (Segmentasi)	0.364

Hasil evaluasi akan diukur menggunakan mAP, yaitu rata-rata dari beberapa AP yang dihitung berdasarkan berbagai kondisi IoU dan ukuran objek, dengan analisis mendalam pada segmentasi dan prediksi box.

D. Implementasi Model pada Gambar



GAMBAR 1 (D)

Hasil prediksi Mask dan objek deteksi: Sapi (a) kambing (b)

Hasil prediksi diterapkan pada beberapa gambar, diikuti dengan analisis mendalam terhadap objek yang berhasil diprediksi dan objek yang berhasil tersegmentasi.

V. KESIMPULAN

Penelitian ini menunjukkan bahwa instance segmentation dengan model Mask R-CNN efektif untuk deteksi objek, menghasilkan akurasi deteksi 94.73% dan segmentasi 90%. Evaluasi menggunakan Average Precision (AP) menunjukkan AP maksimum 62.6% untuk deteksi dan 61.1% untuk segmentasi pada IoU 0.5. Nilai mean Average Precision (mAP) adalah 42.7% untuk deteksi dan 36.4% untuk segmentasi, menunjukkan performa yang baik dalam berbagai kondisi dan lebih unggul dibandingkan penelitian sebelumnya yang hanya menggunakan satu kelas

[1] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in ' Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961– 2969.

[2] Ochoa H, dan Rao K R. 2003. A Hybrid DWT-SVD Image-Coding System (HDWTSVD) for Color Images. Systemics. Cybernetics and Informatics.1:2 64-69

[3] F. A. Kurniadi, C. Setianingsih, and R. E. Syaputra, "Sistem deteksi sapi pada peternakan dari citra dan video uav menggunakan algoritma yolo," eProceeding of Engineering, vol. 10, no. 5, pp. 4582–4589, Oktober 2023

[4] H. M. A. Nayottama, M. Rivai, and H. Pirngadi, "Sistem keamanan pada peternakan sapi menggunakan kamera termal dan metode algoritma yolo," Jurnal Teknik ITS, vol. 12, no. 2, 2023.

[5] H. Zhao, R. Mao, M. Li, B. Li, and M. Wang, "Sheepinst: A high-performance instance segmentation of sheep images based on deep learning,"

[6] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollar, "Microsoft coco: Common objects ' in context," European conference on computer vision, 2014.

[7] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," arXiv preprint arXiv:1511.08458, 2015.

[8] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in ' Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961– 2969.