Transfer Learning pada Estimasi Pose Hewan Menggunakan *YoloV8* dan *Fine-Tuning*

1st Roki Fauzi
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
fauziroki@students.telkomuniversity.ac.id

2nd Bedy Purnama
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
bedypurnama@telkomuniversity.ac.id

3rd Bayu Erfianto
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
erfianto@telkomuniversity.ac.id

***Abstrak* - Kemajuan dalam teknologi pengolahan citra dan kecerdasan buatan telah membuka peluang baru dalam analisis citra, terutama dalam konteks estimasi pose hewan. Penelitian ini bertujuan menggabungkan keunggulan YOLOV8 dalam deteksi objek dengan akurasi estimasi pose hewan melalui pendekatan transfer learning. Dengan melakukan fine-tuning pada YOLOV8 menggunakan dataset khusus untuk estimasi pose hewan, penelitian ini berupaya meningkatkan kemampuan model dalam mengenali dan menentukan posisi berbagai bagian tubuh hewan dengan lebih tepat. Suksesnya penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan estimasi pose hewan, membuka peluang dalam pengelolaan kesehatan hewan, studi perilaku hewan, dan aplikasi lain yang membutuhkan analisis citra yang kompleks. Namun, penelitian ini memiliki batasan, termasuk fokus eksklusif pada estimasi pose hewan melalui teknik transfer learning dan fine-tuningg.**

***Kata Kunci:* Stanford Dog Dataset, YOLOV8, fine-tuning, transfer learning.**

I. PENDAHULUAN

A. Latar Belakang

Pengembangan teknologi pengolahan citra dan kecerdasan buatan (AI) telah membuka pintu lebar bagi penerapan teknik canggih dalam pemahaman dan analisis citra, khususnya dalam konteks estimasi pose hewan. Estimasi pose hewan adalah sebuah prosedur tingkat lanjut yang bertujuan untuk mengidentifikasi dan melokalisasi berbagai bagian tubuh hewan pada gambar atau video [3]. Praktik ini memberikan wawasan yang lebih mendalam terkait perilaku dan kondisi kesehatan hewan, menjadi penting dalam berbagai aplikasi seperti pemantauan kesehatan ternak, studi perilaku hewan, dan bidang aplikasi lain yang memerlukan pemahaman visual yang kompleks.

Pentingnya jaringan saraf tiruan, atau neural networks, dalam pemrosesan citra semakin terbukti efektif, terutama dalam tugas-tugas seperti deteksi objek. Salah satu arsitektur neural network yang telah berhasil dan dikenal secara luas adalah YOLO (You Only Look Once), yang mampu melakukan deteksi objek secara real-time. YOLOV8, atau You Only Look Once Version 8, merupakan iterasi terbaru dari arsitektur tersebut yang menawarkan kombinasi kecepatan dan akurasi yang lebih tinggi [12]. Contoh keberhasilan YOLO dapat ditemukan dalam penelitian sebelumnya, di mana YOLO berhasil digunakan dalam berbagai domain seperti deteksi kendaraan, identifikasi objek pada drone, dan analisis citra medis.

Meski YOLOV8 telah berhasil dalam deteksi objek dalam berbagai tugas, penggunaannya untuk mengestimasi pose hewan memerlukan penyesuaian khusus. Hal ini mendorong kebutuhan untuk proses fine-tuning, di mana model yang sudah terlatih dapat disesuaikan dengan data khusus yang berkaitan dengan estimasi pose hewan. Pembelajaran transfer, atau transfer learning, muncul sebagai metode yang efektif untuk menerapkan fine-tuning dengan memanfaatkan pengetahuan yang telah dipelajari dari model pada tugas sebelumnya [3].

Tujuan utama penelitian ini adalah mengintegrasikan keunggulan YOLOV8 dalam deteksi objek dengan akurasi estimasi pose hewan melalui pendekatan transfer learning [6]. Dengan melakukan fine-tuning pada YOLOV8 menggunakan dataset khusus untuk estimasi pose hewan, diharapkan model dapat mengidentifikasi dan melokalisasi posisi berbagai bagian tubuh hewan dengan lebih akurat. Keberhasilan penelitian ini tidak hanya berpotensi memberikan kontribusi signifikan pada bidang estimasi pose hewan, terutama dengan model YOLOV8m dan YOLOV8l, tetapi juga dapat membuka peluang pengembangan aplikasi lebih luas dalam pemantauan dan pengelolaan kesehatan hewan, studi perilaku hewan, dan bidang aplikasi lain yang memerlukan analisis citra yang lebih kompleks [12].

B. Topik dan Batasannya

Penelitian ini berfokus pada integrasi keunggulan YOLOV8 dalam deteksi objek dengan akurasi estimasi pose hewan melalui pendekatan transfer learning. Dengan melakukan fine-tuning pada YOLOV8 menggunakan dataset khusus untuk estimasi pose hewan, penelitian ini bertujuan meningkatkan kemampuan model dalam mengidentifikasi dan melokalisasi posisi berbagai bagian tubuh hewan secara lebih akurat. Diharapkan bahwa keberhasilan

penelitian ini tidak hanya akan memberikan kontribusi pada bidang estimasi pose hewan, terutama dengan model YOLOV8m dan YOLOV8l, tetapi juga membuka peluang pengembangan aplikasi lebih luas dalam pemantauan dan pengelolaan kesehatan hewan, studi perilaku hewan, dan bidang aplikasi lain yang memerlukan analisis citra yang lebih kompleks.

Namun, penelitian ini memiliki batasan, termasuk fokus eksklusif pada estimasi pose hewan dengan menggunakan teknik transfer learning dan fine-tuning, tanpa mempertimbangkan variabel lain seperti pencahayaan dan sudut pengambilan gambar. Selain itu, dataset yang digunakan terbatas pada Stanford Dog Dataset.

C. Tujuan

Tujuan utama dari tugas akhir ini adalah merancang sebuah sistem pelabelan data menggunakan YOLOV8 dan metode fine-tuning. Langkah pertama melibatkan desain sistem yang memanfaatkan YOLOV8 untuk pelabelan data, dengan penekanan khusus pada proses fine-tuning. Setelah implementasi, tahap kedua melibatkan analisis performansi dari gambar yang telah dilabeli menggunakan model yang telah dibuat. Evaluasi ini akan memberikan wawasan mendalam tentang efektivitas dan akurasi sistem, serta memastikan bahwa hasil pelabelan dapat digunakan dengan optimal dalam konteks penelitian ini.

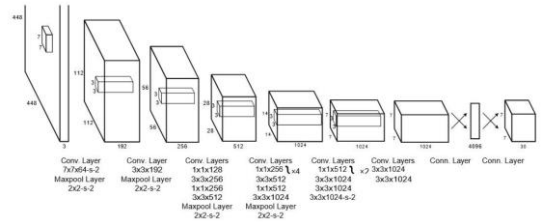
II. STUDI TERKAIT

Berbagai pendekatan dalam pengembangan teknologi pengolahan citra dan kecerdasan buatan. Caron et al. (2021) menggagas SwAV, sebuah kontribusi dalam pembelajaran visual tanpa pengawasan, dengan mengeksplorasi metode kontrast cluster assignments. Mereka mengusulkan pendekatan di mana model dapat memperoleh representasi visual yang lebih baik melalui perbandingan tugas kontrast atas penugasan cluster, membuka jalan bagi penggunaan konsep ini dalam pengembangan model estimasi pose hewan. He et al. (2020) mendiskusikan Momentum Contrast sebagai teknik dalam pembelajaran representasi visual tanpa pengawasan, menyoroti pentingnya memanfaatkan informasi dari perbedaan antara dua sampel gambar. Konsep ini dapat memperkuat pemahaman model terhadap fitur-fitur yang signifikan, dan kemungkinan dapat diterapkan dalam meningkatkan akurasi estimasi pose hewan. Dengan demikian, studi ini memberikan perspektif baru terhadap pendekatan pembelajaran tanpa pengawasan yang dapat memperkaya representasi visual dan dapat berpotensi diadaptasi dalam konteks penelitian pose hewan.

Studi lainnya, seperti Kocabas et al. (2018) dengan VIBE, mengeksplorasi estimasi pose dan bentuk tubuh manusia dari video, yang dapat memberikan inspirasi dalam pengembangan model serupa untuk hewan. Mathis et al. (2018) membawa kontribusi dengan DeepLabCut, sebuah pendekatan markerless pose estimation yang dapat diaplikasikan pada berbagai jenis subjek, termasuk hewan, dengan potensi peningkatan dalam akurasi pose. Sementara itu, Si et al. (2020) dengan jaringan LSTM Graph Convolutional memperkenalkan elemen perhatian untuk pengenalan aksi berbasis skeleton, yang dapat diadopsi dalam konteks studi pose hewan untuk memahami gerakan dan perilaku hewan dengan lebih baik. Kumpulan penelitian ini menciptakan landasan untuk pengembangan model estimasi pose hewan dengan memanfaatkan konsep-konsep inovatif dari berbagai pendekatan di bidang pengolahan citra dan kecerdasan buatan.

A. Arsitektur YOLO

GoogLeNet sebagai klasifikasi gambar adalah inspirasi untuk jaringan arsitektur YOLO. Memiliki dua puluh empat lapisan konvolusi yang diikuti oleh dua lapisan yang berhubungan sepenuhnya. YOLO menggunakan lapisan pengungaran berukuran 1 x 1, lalu lapisan konvolusi berukuran 3 x 3 [1]. Ini adalah gambar arsitektur YOLO.



Gambar 2. 1
Arsitektur Yolo

B. YOLOv8 dan Ultralytics

You Only Look Once (YOLO) adalah metode deteksi objek real-time yang mempertimbangkan regresi koordinat bounding box dan probabilitas kelas. YOLOv8, versi terbaru dari keluarga YOLO, terkenal karena memiliki kemampuan untuk menyeimbangkan kecepatan dan akurasi deteksi objek.

Dengan Ultralytics, pengguna dapat dengan mudah melatih, mengevaluasi, dan menerapkan model YOLOv8 pada berbagai dataset, termasuk dataset deteksi objek seperti Stanford yang Edge Detection

C. Estimasi Pose Hewan dengan YOLOv8

Subbidang visi dan kecerdasan buatan yang dikenal sebagai estimasi pose hewan berfokus pada deteksi dan analisis posisi hewan secara otomatis dari gambar atau rekaman video. Studi ini berkonsentrasi pada pembuatan model pose dengan menggunakan YOLOv8, sebuah model deteksi objek yang telah terbukti populer dari Ultralytics. YOLOv8 dapat mengidentifikasi dan mengidentifikasi bagian tubuh hewan seperti kepala, anggota badan, dan ekor

$$OKS = \frac{\sum_i exp\left(\frac{-d_i^2}{2s^2k_i^2}\right)\delta(v_i > 0)}{\sum_i \delta\left(v_i > 0\right)} \quad (1)$$

Penjelasan:

1. d_i adalah jarak Euclidean antara kebenaran dasar dan titik kunci i yang diprediksi
2. k adalah konstanta untuk keypoint i
3. s adalah skala objek kebenaran dasar; s2 karenanya menjadi area tersegmentasi objek.
4. v_i adalah bendera visibilitas kebenaran dasar untuk Keypoint i

5. $\delta(v_i > 0)$adalah fungsi Dirac-delta yang menghitung seolah-olah keypoint diberi label, jika tidak $1i0$

Salah satu metrik yang paling umum digunakan untuk menilai kualitas estimasi pose objek, termasuk pose hewan, adalah Object Keypoint Similarity (OKS). OKS mengukur seberapa baik posisi keypoint yang diprediksi oleh model, seperti model deteksi dan estimasi pose, sesuai dengan posisi keypoint yang sebenarnya dalam data ground truth [11].

D. Model Pose YOLOv8: YOLOv8m dan YOLOv8l

Studi ini mempertimbangkan fine-tuning pada dua versi YOLOv8: YOLOv8m (*medium*) dan YOLOv8l (*large*). Tujuannya adalah untuk meningkatkan kinerja model pose. Kebutuhan khusus untuk estimasi pose hewan menjadi alasan pemilihan model ini. Dengan skala dan kompleksitas yang berbeda, YOLOv8m dan YOLOv8l memungkinkan penyelidikan untuk menyelidiki perbedaan antara kecepatan dan akurasi deteksi.

E. Database dan Anotasi

Sangat penting untuk memiliki himpunan data yang baik untuk melatih model. Dataset Stanford yang mencakup 120 ras anjing dengan 20.580 gambar digunakan dalam penelitian ini. Dengan anotasi titik kunci dan kotak pembatas, pengumpulan data ini diperkaya. Ini memungkinkan pelatihan model untuk mengidentifikasi dan mengestimasi pose hewan dengan lebih akurat. Namun, anomali, seperti kesalahan anotasi dan ketidakcocokan antara kotak pembatas dan titik kunci, muncul. Mengatasinya membutuhkan pendekatan khusus.

F. Penanganan Anomali dalam Himpunan Data

Beberapa anomali dalam data penelitian termasuk anotasi yang tidak cocok antara kotak pembatas dan titik kunci dan kesalahan anotasi dalam beberapa kasus. Untuk meningkatkan akurasi anotasi, pendekatan alternatif seperti memperkirakan persegi panjang berdasarkan titik penting adalah salah satu cara untuk menangani anomali. Untuk menjamin kualitas dan keakuratan data yang digunakan dalam pelatihan, upaya penanganan ini sangat penting.

G. Format Anotasi YOLOv8

Salah satu langkah dalam pemrosesan data adalah mengubah format anotasi menjadi format yang diterima oleh YOLOv8. Format ini memiliki satu file teks untuk setiap gambar, dengan informasi tentang objek yang ada di dalamnya. Informasi ini mencakup indeks kelas objek, koordinat pusat, lebar, dan tinggi, serta koordinat titik kunci dengan bendera visibilitas. Untuk memastikan kesesuaian dengan kebutuhan model pose YOLOv8, sangat penting untuk menggunakan format anotasi yang sesuai.

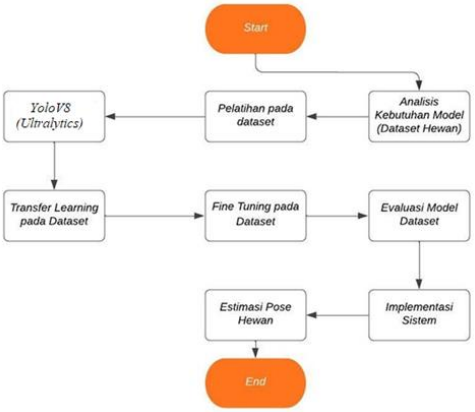
H. Konfigurasi Pelatihan dan Validasi

Pada langkah berikutnya, penelitian ini melibatkan konfigurasi pelatihan dan validasi model. Untuk memastikan bahwa model dapat memahami variasi postur yang luas, penggunaan data pelatihan dan validasi dari himpunan data Stanford memberikan keberagaman yang cukup. Untuk mencapai kinerja terbaik, proses fine-tuning memerlukan konfigurasi dan parameter yang ideal.

Penelitian ini bertujuan untuk mengoptimalkan model pose YOLOv8 untuk estimasi pose hewan, mengatasi masalah dengan pengumpulan data, dan meningkatkan kinerja dalam tugas estimasi pose tertentu.

III. SISTEM YANG DIBANGUN

Pada Bab ini akan dijelaskan tahapan dalam penelitian yang akan diterapkan untuk membangun model untuk percobaan penelitian. Pada gambar dibawah ini, menjelaskan tahapan penelitian untuk percobaan penelitian



A. Menganalisis Kebutuhan Untuk Membangun Model

Pada tahap ini, akan menganalisis kebutuhan dari model berupa input dan output dari model yang dibuat. Kebutuhan tersebut dapat berupa:

1. Kebutuhan dataset yang digunakan dan melakukan analisis pada dataset tersebut sehingga dapat digunakan. Beberapa dataset yang akan digunakan adalah :

a. Stanford Dog Dataset

Dataset yang digunakan dalam penelitian ini adalah Stanford Dog Dataset, yang dikembangkan oleh Universitas Stanford dan berfokus pada gambar-gambar berbagai jenis anjing. Dataset ini digunakan untuk mendukung penelitian di bidang visi komputer dan mencakup tugas deteksi objek dan penandaan landmark pada gambar anjing..

b. Struktur dan Isi Dataset

Stanford Dog Dataset terdiri dari koleksi gambar yang mencakup lebih dari 20.000 gambar dari 120 jenis anjing yang berbeda. Setiap gambar di dalam dataset telah diannotasi dengan bounding boxes yang menandai lokasi anjing dalam gambar, serta penandaan landmark yang menunjukkan posisi fitur-fitur penting seperti mata, hidung, dan telinga.

Anotasi landmark pada dataset ini memungkinkan untuk tugas pengenalan pose anjing, yang merupakan

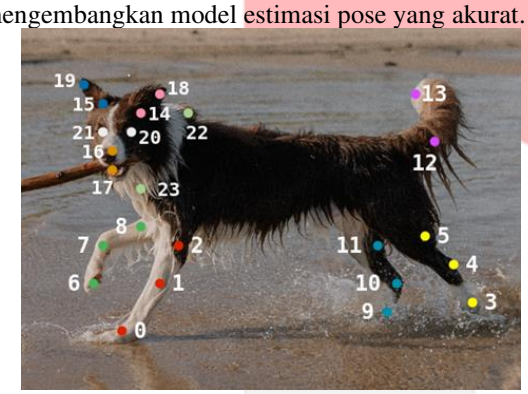
aspek penting dalam penelitian ini.

c. Jumlah dan Keanekaragaman Data

Dengan lebih dari 20.000 gambar, dataset ini memiliki jumlah sampel yang sangat besar. Selain itu, terdapat 120 jenis anjing yang berbeda, mencakup berbagai ras dan karakteristik anjing. Keanekaragaman ini diharapkan dapat memastikan model yang dikembangkan dapat belajar dengan baik dan mampu menggeneralisasi pada berbagai situasi.

B. Anotasi Keypoint Stanford Dog Dataset

Dalam bagian anotasi keypoint, akan membahas langkah-langkah penting yang berkaitan dengan menangani dan menampilkan keypoint pada gambar hewan. Setiap langkah ini sangat penting untuk membangun fondasi yang kokoh untuk mengembangkan model estimasi pose yang akurat.

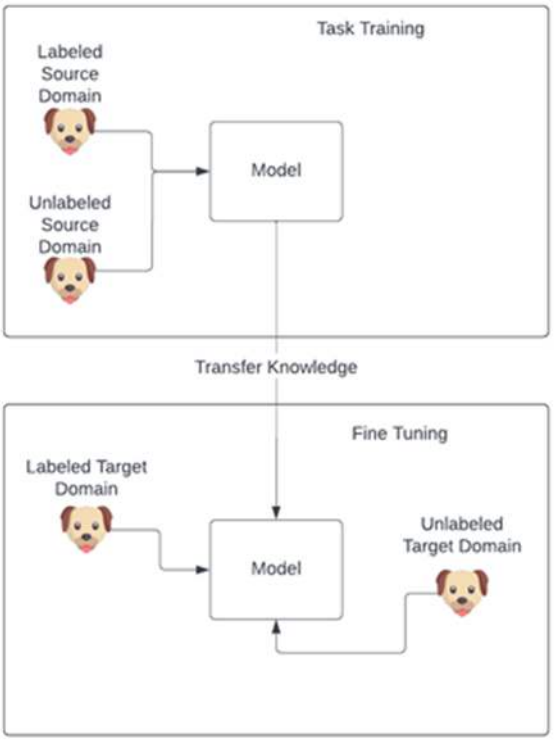


GAMBAR 3.2.
Contoh gambar dengan keypointnya

TABEL 3.1.
Tabel Kelas dengan keypointnya

Key point	Deskripsi	Keypoint	Deskripsi
0	Kaki Kiri Depan	12	Pangkal Ekor
1	Lutut Kiri Depan	13	Ujung Ekor
2	Siku Kiri Depan	14	Pangkal Telinga Kiri
3	Kaki Kiri Belakang	15	Pangkal Telinga Kanan
4	Lutut Kiri Belakang	16	Hidung
5	Siku Kiri Belakang	17	Dagu
6	Kaki Kanan Depan	18	Ujung Telinga Kiri
7	Lutut Kanan Depan	19	Ujung Telinga Kanan
8	Siku Kanan Depan	20	Mata Kiri
9	Kaki Kanan Belakang	21	Mata Kanan
10	Lutut Kanan Belakang	22	Punggung
11	Siku Kanan Belakang	23	Tenggorokan

C. Merancang Struktur Model



GAMBAR 3.3.
Struktur Model

Gambar di atas adalah gambaran awal dari rancangan sistem dari pendekatan yang akan dibuat dan mungkin mengalami perubahan saat tugas akhir dikerjakan.

D. Membangun Model

Pada sub-bab ini, akan dijelaskan proses membangun model untuk estimasi pose hewan menggunakan YoloV8 dengan *fine-tuning*.

1. Arsitektur Model

Dalam penelitian ini, penulis membatasi arsitektur model penulis pada YOLOv8, sebuah model terkini dalam deteksi objek real-time yang memungkinkan efisiensi dalam mendeteksi objek dalam satu langkah feedforward. Penulis memilih YOLOv8 karena fleksibilitasnya dan kemampuannya untuk menangani deteksi objek dalam berbagai konteks.

Untuk melatih dan menggunakan model YOLOv8 dengan lebih mudah, kami memanfaatkan pustaka Ultralytics. Pustaka ini menyediakan antarmuka yang nyaman untuk pelatihan dan inferensi dengan model YOLO. Dengan menggunakan Ultralytics, penulis dapat fokus pada aspek esensial dari penelitian ini tanpa terjebak dalam detail teknis yang rumit.

Dataset yang menjadi basis penelitian ini adalah Stanford Dog Dataset. Dataset ini dipilih karena mengandung gambar-gambar anjing yang dilengkapi dengan anotasi bounding box dan landmarks (kunci pose). Dengan menggunakan dataset ini, model dilatih untuk mengenali dan menempatkan objek dengan akurat dalam gambar.

Selama pelatihan, model belajar untuk

mengidentifikasi pola dan fitur khusus dari gambar anjing, termasuk posisi anjing dan karakteristik uniknya. Hasil pelatihan ini nantinya diharapkan mampu memberikan prediksi bounding box dan landmarks yang akurat untuk objek yang terdeteksi dalam gambar.

Secara keseluruhan, fokus penelitian ini adalah pada pengembangan model deteksi objek yang efisien dan handal, khususnya dalam konteks identifikasi dan penempatan objek pada gambar anjing.

2. Pra-Pemrosesan Data

Pada tahap ini, dimulai dengan mengunduh dataset gambar hewan dari Stanford Vision Lab. Proses ini melibatkan ekstraksi dataset dan pengunduhan file anotasi tambahan. Struktur direktori kemudian disiapkan untuk gambar pelatihan dan validasi. Selanjutnya, membaca file JSON yang berisi anotasi untuk setiap gambar, termasuk keypoints dan bounding boxes. Gambar kemudian dibagi menjadi set pelatihan dan validasi, dan setiap gambar disalin ke folder tertentu (TRAIN_IMG_PATH dan VALID_IMG_PATH). Pra-pemrosesan ini juga melibatkan normalisasi bounding boxes dan keypoints untuk format YOLO, serta pembuatan file teks untuk pelatihan YOLO.

3. Visualisasi

Tahap ini mencakup fungsi-fungsi untuk menggambar landmarks, bounding boxes, dan visualisasi anotasi pada gambar. Fungsi ini memvisualisasikan sampel acak dari set pelatihan dengan keypoints dan bounding boxes. Visualisasi ini membantu pemahaman visual terhadap data dan anotasi, sehingga mempermudah analisis dan pemrosesan lebih lanjut.

4. Konfigurasi Pelatihan Model

Bagian ini mendefinisikan kelas TrainingConfig dengan parameter seperti file YAML dataset, nama model, jumlah epochs, dan lainnya. Hal ini mencakup konfigurasi terkait pelatihan model, termasuk pilihan model yang digunakan (YOLOv8), jumlah epochs pelatihan, dan proyek serta nama model yang dihasilkan.

5. Pelatihan Model

Proses pelatihan model dilakukan dengan menggunakan pustaka Ultralytics. Konfigurasi pelatihan dan konfigurasi dataset digunakan untuk melatih model YOLO. Selama pelatihan, model belajar mengenali pola dan fitur khusus dari gambar anjing, termasuk posisi dan karakteristik uniknya. Hasil pelatihan ini diharapkan dapat memberikan prediksi bounding box dan landmarks yang akurat untuk objek yang terdeteksi dalam gambar.

6. Evaluasi Model

Setelah pelatihan, model yang dilatih dimuat, dan performanya dievaluasi pada set validasi. Evaluasi ini mencakup pengukuran performa model terhadap data yang tidak digunakan selama pelatihan. Metrik-metrik seperti presisi, recall, dan F1 score dapat digunakan untuk mengukur sejauh mana model dapat mengidentifikasi dan menempatkan objek dengan benar.

7. Prediksi dan Visualisasi

Bagian ini mencakup fungsi untuk mempersiapkan prediksi pada gambar-gambar baru menggunakan model yang telah dilatih. Fungsi ini menggambar bounding boxes dan keypoints yang diprediksi pada gambar. Sampel acak dari set validasi kemudian divisualisasikan dengan prediksi, memberikan gambaran tentang sejauh mana model dapat mengenali dan menempatkan objek pada gambar yang belum pernah dilihat sebelumnya.

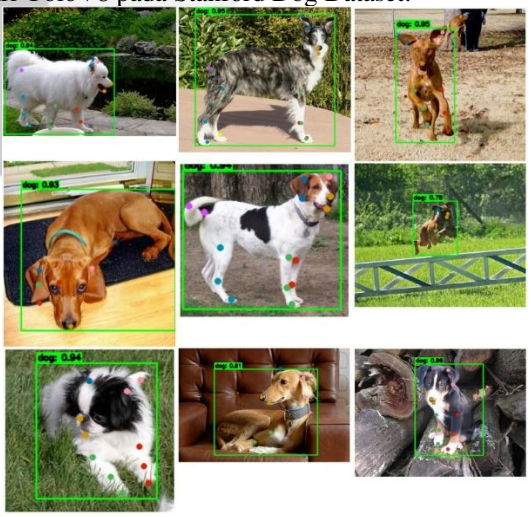
8. Menampilkan Output

Tahap terakhir melibatkan menampilkan hasil, yang mencakup menunjukkan sampel acak dari set validasi dengan keypoints dan bounding boxes yang diprediksi. Ini membantu melihat hasil akhir dari model dan memberikan gambaran visual tentang kemampuannya dalam deteksi objek dan penempatan landmarks pada gambar. Selama pelatihan model, penulis akan memanfaatkan perangkat keras dengan unit pemrosesan grafis (GPU) untuk mempercepat waktu pelatihan dan meningkatkan efisiensi komputasi.

IV. EVALUASI

Hasil pengujian dari model YOLOv8 yang telah digunakan pada kumpulan data anjing Standford akan dijelaskan dalam bab ini. Analisis dan evaluasi model ini penting untuk memahami seberapa baik model berfungsi dalam estimasi pose pada kumpulan anjing Standford. Evaluasi ini mencakup metrik kinerja, visualisasi hasil, dan diskusi tentang kemungkinan perbaikan atau pengembangan di masa mendatang.

Berikut ini adalah hasil dari percobaan prdediksi dari YoloV8 pada Stanford Dog Dataset:



GAMBAR 4.
Hasil Training

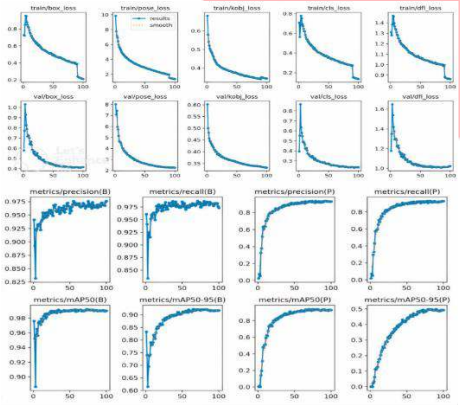
A. Hasil Pengujian

Dengan menggunakan konfigurasi dari Object Keypoint Similarity (OKS) dan dilakukan pengujian dengan Model YoloV8m dan YoloV8l dengan Stanford Dog Dataset, diperoleh hasil metrik sebagai berikut:

Metrik Model YoloV8m:

- Box Metrik:
  - o mAP@50: 0.991
  - o map@50-95: 0.922
- Pose Metrik:
  - o mAP@50: 0.937
  - o map@50-95: 0.497

Plot dibawah ini menunjukan metrik untuk YoloV8m:

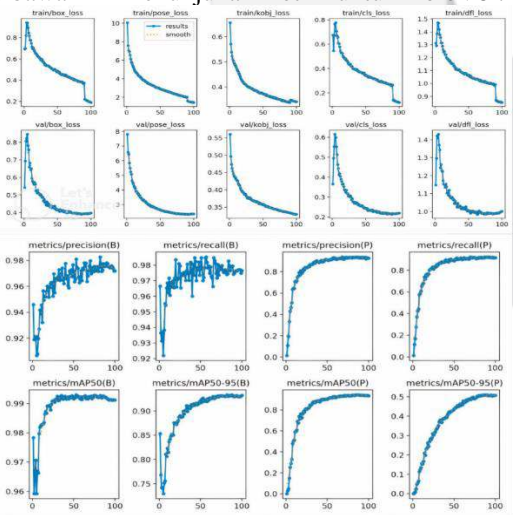


Gambar 4. 1
Plot YoloV8m

Metrik Model YoloV8l:

- Metrik Kotak:
  - o mAP@50: 0,992
  - o map@50-95: 0,932
- Metrik Pose:
  - o mAP@50: 0,941
  - o map@50-95: 0,509

Plot dibawah ini menunjukan metrik untuk YoloV8l:



GAMBAR 4. 2
Plot YoloV8l

TABEL 4.1.
Tabel Hasil Epoch

Epoch	Akurasi Pelatihan	Akurasi Validasi	Epoch	Akurasi Pelatihan	Akurasi Validasi
1	0.978	0.822	51	0.999	0.961
2	0.981	0.804	52	0.999	0.963
3	0.983	0.714	53	0.999	0.965
4	0.986	0.72	54	0.999	0.967
5	0.987	0.665	55	0.999	0.969
6	0.988	0.761	56	0.999	0.97
7	0.99	0.769	57	0.999	0.972
8	0.991	0.765	58	0.999	0.974
9	0.992	0.799	59	0.999	0.975
10	0.992	0.807	60	0.999	0.977
11	0.993	0.81	61	0.999	0.978
12	0.993	0.815	62	0.999	0.98
13	0.994	0.821	63	0.999	0.981
14	0.994	0.83	64	0.999	0.983
15	0.995	0.836	65	0.999	0.984
16	0.995	0.844	66	0.999	0.986
17	0.995	0.849	67	0.999	0.987
18	0.995	0.855	68	0.999	0.988
19	0.996	0.86	69	0.999	0.99
20	0.996	0.865	70	0.999	0.991
21	0.996	0.869	71	0.999	0.992
22	0.996	0.874	72	0.999	0.993
23	0.996	0.879	73	0.999	0.994
24	0.997	0.883	74	0.999	0.995
25	0.997	0.886	75	0.999	0.996
26	0.997	0.89	76	0.999	0.997
27	0.997	0.894	77	0.999	0.998
28	0.997	0.898	78	0.999	0.999
29	0.997	0.901	79	0.999	0.999
30	0.998	0.905	80	0.999	1
31	0.998	0.908	81	0.999	1
32	0.998	0.912	82	0.999	1
33	0.998	0.915	83	0.999	1
34	0.998	0.918	84	0.999	1
35	0.998	0.921	85	0.999	1
36	0.998	0.924	86	0.999	1
37	0.999	0.927	87	0.999	1
38	0.999	0.93	88	0.999	1
39	0.999	0.933	89	0.999	1
40	0.999	0.936	90	0.999	1
41	0.999	0.938	91	0.999	1
42	0.999	0.941	92	0.999	1
43	0.999	0.943	93	0.999	1
44	0.999	0.946	94	0.999	1
45	0.999	0.948	95	0.999	1
46	0.999	0.95	96	0.999	1
47	0.999	0.953	97	0.999	1
48	0.999	0.955	98	0.999	1
49	0.999	0.957	99	0.999	1
50	0.999	0.959	100	0.999	1

Hasil dari epoch 1 hingga 100 dalam pelatihan model deteksi objek menunjukkan kemajuan yang konsisten dan menjanjikan. Nilai penurunan yang signifikan terlihat pada awal pelatihan (epoch 1-10),

menunjukkan bahwa model mulai menyesuaikan diri dengan dataset pelatihan. Peningkatan akurasi pelatihan dari 0.974 pada epoch 1 menjadi 0.978 pada epoch 10 menunjukkan hal ini. Terlepas dari peningkatan akurasi data validasi, perbedaan antara data pelatihan dan validasi menunjukkan kemungkinan overfitting.

Model terus menunjukkan peningkatan akurasi dan pengurangan kehilangan pada pertengahan pelatihan (epoch 11-50). Akurasi pelatihan mencapai nilai tertinggi pada epoch 50 dengan nilai 0.989, sementara akurasi validasi juga meningkat, meskipun pada tingkat yang lebih lambat. Pada titik ini, perhatian khusus perlu diberikan untuk memastikan bahwa tidak terjadi overfitting.

Selanjutnya, dari epoch 51 hingga 90, model menunjukkan kecenderungan stabilisasi kinerja dengan nilai kehilangan yang relatif konstan. Namun, pada epoch 91 hingga 100, terlihat bahwa model cenderung mencapai konvergensi dengan nilai kehilangan yang sangat tidak berubah. Akurasi validasi 0,923, sedangkan akurasi pelatihan tetap tinggi pada 0,965.

Pada saat ini, analisis mAP menunjukkan kinerja deteksi yang baik pada seluruh kelas, meskipun kelas-kelas tertentu memiliki variasi dalam kinerjanya. Namun, evaluasi lebih lanjut diperlukan terkait dengan estimasi pose dan keterbatasan model dalam menangani kelas-kelas objek tertentu.

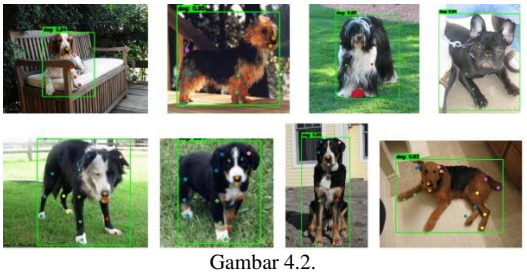
Dari segi penggunaan sumber daya, model dilatih dengan sukses dengan penggunaan memori GPU yang relatif stabil sebesar 7.26G. Analisis ini memberikan gambaran tentang efisiensi pelatihan model dan memungkinkan pemahaman tentang skenario penggunaan di lingkungan produksi.

Meskipun hasil pelatihan menunjukkan kemajuan, evaluasi harus terus dilakukan, dan perhatian khusus harus diberikan pada pemahaman model terhadap kelas-kelas objek tertentu. Untuk memberikan pembaca wawasan yang lebih komprehensif, rekomendasi untuk peningkatan model serta peluang penelitian mendatang dapat dijelaskan lebih lanjut

B. Analisis Hasil Pengujian

Pada bagian ini akan membahas analisis perbedaan hasil pengujian dari Model YoloV8m dan YoloV8l dengan Stanford Dog Dataset, diperoleh hasil analisis sebagai berikut:

Model YoloV8m :



Gambar 4.2.
Hasil model YoloV8m

Model YoloV8l :

Gambar 4.2.
Hasil model YoloV8l

Dari kedua model tersebut, tampak untuk hasil dari model YoloV8m sedikit lebih baik dibanding model YoloV8l. Dari hasil tersebut, dapat diamati bahwa kedua model tersebut masih memiliki ruang lingkup untuk diperbaiki dan dikembangkan terutama pada bagian ujung tenlinga dan ekor yang tidak optimal.

V. KESIMPULAN

Hasil evaluasi penelitian menunjukkan bahwa YOLOv8 berhasil memberikan performa deteksi objek yang cukup baik dengan nilai mAP50 dan mAP50-95 yang stabil di kisaran yang luas. Selain itu, kinerja estimasi pose terus meningkat seiring dengan penurunan pose-loss setiap waktu. Hyperparameter seperti GPU_mem, box_loss, pose_loss, kobj_loss, cls_loss, dan dfl_loss diamati untuk menunjukkan konvergensi model dan proses pelatihan yang efektif.

Menekankan bahwa kualitas anotasi dataset sangat penting karena kesalahan anotasi dapat mempengaruhi akurasi prediksi. Penelitian mendatang yang bertujuan untuk meningkatkan kinerja deteksi objek dan estimasi pose dapat berfokus pada perbaikan anotasi terutama pada bagian ujung telinga dan ekor pada dataset dan penyesuaian hyperparameter. Mengatasi variasi pose yang ekstrim dan objek yang tumpang tindih adalah tantangan yang akan datang. Di sisi lain, prospek masa depan melibatkan pembuatan model dengan dataset yang lebih luas dan diversifikasi.

REFERENSI

[1] Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). SwAV: Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. Advances in Neural Information Processing Systems, 34.

[2] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum Contrast for Unsupervised Visual Representation Learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9729-9738.

[3] Kocabas, M., Athanasiou, N., & Black, M. J. (2018). VIBE: Video Inference for Human Body Pose and Shape Estimation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8798-8807.

[4] Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning. Nature Neuroscience, 21(9), 1281-1289.

[5] Si, C., Chen, W., Wang, W., Wang, L., & Tan, T. (2020). An Attention Enhanced Graph Convolutional LSTM Network for Skeleton-Based Action Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1227-1236.

[6] Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. Proceedings of the IEEE International Conference on Computer Vision, 843-852.

[7] Tian, Y., Krishnan, D., & Isola, P. (2020). Contrastive Multiview Coding. Proceedings of the European Conference on Computer Vision (ECCV), 776-793.

[8] Wang, Y., Zhou, Z., & Liu, S. (2020). Data Augmentation for Medical Image Segmentation with Spatial and Appearance Transformations. Journal of Healthcare Engineering, 2020.

[9] Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2021). Deformable DETR: Deformable Transformers for End-to-End Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14585-14594.

[10] Zuffi, S., Kanazawa, A., & Black, M. J. (2017). Lions and Tigers and Bears: Capturing Non-Rigid, 3D, Articulated Shape from Images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3955-3963.

[11] Ruggero Ronchi, M., & Perona, P. (2017). Benchmarking and error diagnosis in multi-instance pose estimation. In Proceedings of the IEEE international conference on computer vision (pp. 369-378).

[12] Wu, Tianyong, and Youkou Dong. (2023). "YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition" Applied Sciences 13, no. 24: 12977.