

# Deteksi Objek pada Citra Menggunakan Model YOLO

Intara Pratama Harahap<sup>a1</sup>, Agus Muliantara<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>redintara123@gmail.com  
<sup>2</sup>muliantara@unud.ac.id

## Abstract

*Object detection is a crucial task in the field of computer vision and digital image processing, with numerous practical applications. This paper focuses on the implementation of the You Only Look Once (YOLO) model, a deep learning-based approach for object detection. The YOLO model offers several advantages over previous methods, such as simultaneous prediction of bounding boxes and object class probabilities, a relatively simple Convolutional Neural Network (CNN) architecture, and high computational speed, making it suitable for real-time applications. The study utilizes a dataset of 770 images, with 524 for training, 136 for validation, and 110 for testing, specifically focused on detecting various pet animals. The training process involves annotation of the image data, followed by training and validation of the YOLO model. The results demonstrate the model's ability to effectively detect and classify objects, achieving high performance metrics such as precision, recall, and mean Average Precision (mAP) nearing 0.8 towards the end of the training process. Additionally, a confusion matrix is presented, highlighting the model's accuracy in classifying different classes, with the highest accuracy for the 'Cat' class at 95%. The paper concludes by discussing the model's performance and potential areas for improvement.*

**Keywords:** YOLO, You Only Look Once, Citra, Object Detection,

## 1. Pendahuluan

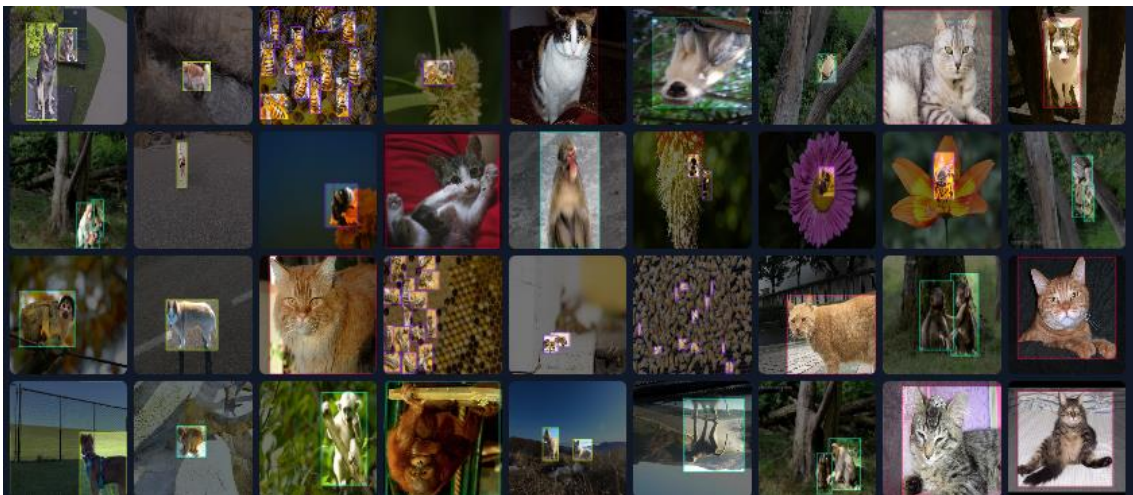
Artificial Intelligence (AI) atau Kecerdasan Buatan adalah ilmu yang dikembangkan dengan kombinasi banyak subyek. Dalam istilah sederhana, kecerdasan buatan adalah memberi mesin dengan kecerdasan manusia, mensimulasikan pemikiran manusia sehingga membantu orang memecahkan masalah dan untuk mewujudkan aplikasi yang lebih canggih seperti produksi dan kehidupan manusia yang dibantu computer. Kecerdasan buatan, cabang ilmu komputer, dianggap sebagai salah satu dari tiga teknologi yang paling maju (Rekayasa genetika, ilmu Nano dan kecerdasan Buatan) pada abad ke-21. Kecerdasan buatan (Artificial Intelligence) memodifikasi atau mengadaptasi computer dari tindakan manusia (termasuk prediksi atau control robot), sehingga bisa menjadi lebih akurat. Kecerdasan buatan telah mencapai perkembangan yang sangat cepat di 30 tahun terakhir dan telah banyak digunakan di banyak negara bidang akademik. Supaya computer dapat bertindak atau serupa dengan manusia, maka komputer harus diberi bekal pengetahuan, dan mempunyai kemampuan untuk menalar. Penerapan AI beraneka ragam. Salah satu kemampuan ai saat ini adalah mengenali objek-objek pada sebuah gambar atau video, singkatnya dapat mendeteksi objek. Deteksi objek merupakan salah satu tugas penting dalam bidang visi komputer dan pengolahan citra digital. Tujuan utama dari deteksi objek adalah untuk mengidentifikasi dan melokalisasi objek-objek spesifik dalam sebuah citra atau video. Deteksi objek memiliki banyak aplikasi praktis, seperti dalam sistem keamanan, kendaraan otonom, sistem pengawasan, analisis citra medis, dan banyak lagi [1]. Dalam beberapa tahun terakhir, pendekatan berbasis deep learning, khususnya Convolutional Neural Network (CNN), telah menjadi metode yang paling populer dan efektif untuk tugas deteksi objek. Salah satu model deteksi objek berbasis deep learning yang telah mencapai popularitas adalah YOLO (You Only Look Once), yang diusulkan oleh Redmon et al. pada tahun 2016 [1]. Model YOLO menawarkan beberapa keunggulan dibandingkan metode deteksi objek sebelumnya, seperti R-CNN [2] dan Faster R-CNN [3]. Pertama, YOLO memprediksi bounding box dan

kemungkinan kelas objek secara simultan, sehingga lebih efisien dalam hal waktu komputasi. Kedua, YOLO menggunakan arsitektur CNN yang relatif sederhana, sehingga lebih mudah diimplementasikan dan dioptimalkan. Ketiga, YOLO memiliki kecepatan yang tinggi, sehingga cocok untuk aplikasi real-time seperti deteksi objek pada video atau kamera yang bergerak. Meskipun demikian, model YOLO juga memiliki beberapa keterbatasan, seperti akurasi yang sedikit lebih rendah dibandingkan metode lain pada objek kecil atau objek dengan variasi bentuk yang ekstrem. Oleh karena itu, pengembangan dan penyempurnaan model YOLO terus dilakukan, seperti versi YOLOv3 yang memiliki peningkatan dalam hal akurasi dan kemampuan mendeteksi objek kecil.

## 2. Metode Penelitian

### 2.1. Dataset

Dataset yang digunakan diambil dari Roboflow. Dataset yang kali ini saya gunakan berupa dataset tentang hewan khususnya hewan peliharaan. Dataset yang totalnya berjumlah 770 gambar. Dataset untuk train berjumlah 524, untuk validasi berjumlah 136 dan terakhir untuk testing berjumlah 110 gambar.



Gambar 1. Contoh Dataset

### 2.2. CNN (Convolutional Neural Network)

Convolutional Neural Network (CNN) adalah salah satu jenis Neural Network (NN) yang biasa digunakan pada data image. CNN bisa digunakan untuk mendeteksi dan mengenali object pada sebuah image. Secara garis besar CNN tidak jauh beda dengan Neural Network biasanya. CNN terdiri dari neuron yang memiliki weight, bias dan activation function. Yang membedakan antara CNN dengan NN ialah arsitektur dari CNN dibagi menjadi dua bagian besar, yaitu Feature Extraction Layer dan Fully Connected Layer [2].

### 2.3. You Only Look Once (YOLO)

Arsitektur YOLO terdiri atas 24 lapisan konvolusional (convolutional layer) dengan 4 lapisan max pooling, yang diikuti oleh 2 lapisan yang terhubung penuh (fully connected layer). Beberapa lapisan konvolusi menggunakan lapisan reduksi 1x1 sebagai alternatif untuk mengurangi kedalaman feature maps [1].

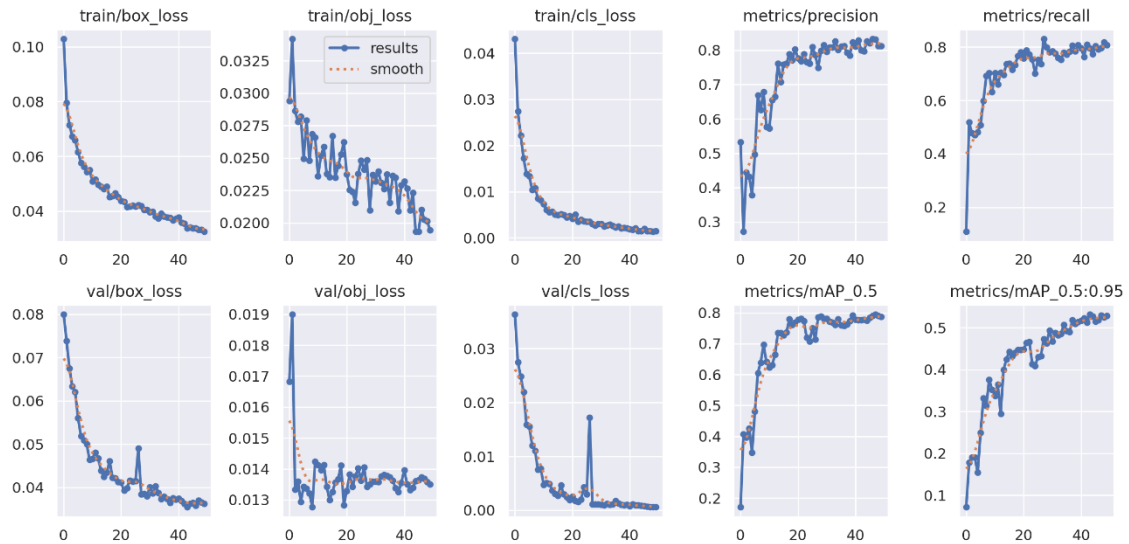
Arsitektur YOLO sesungguhnya cukup sederhana. Sistem akan menerima input citra dengan bentuk (448, 448, 3) yaitu citra berukuran 448 x 448 dengan 3 channel, yang kemudian akan melewati satu kali proses convolutional network hingga menghasilkan output dengan bentuk (7, 7, 30), dimana 7 x 7 merupakan ukuran grid sel ( $S = 7$ ) dan 30 merupakan nilai dari jumlah kotak



dilakukan setelah proses anotasi pada seluruh citra selesai, sehingga diperoleh kumpulan file .txt dari tiap citra yang berisi representasi hasil anotasi citra.

### 3.2. Training dan Validasi Model YOLO V

Proses training model dilakukan pada data training. Metriks evaluasi seperti Precision, Recall, dan mAP diperoleh melalui proses validasi pada model yang dihasilkan.

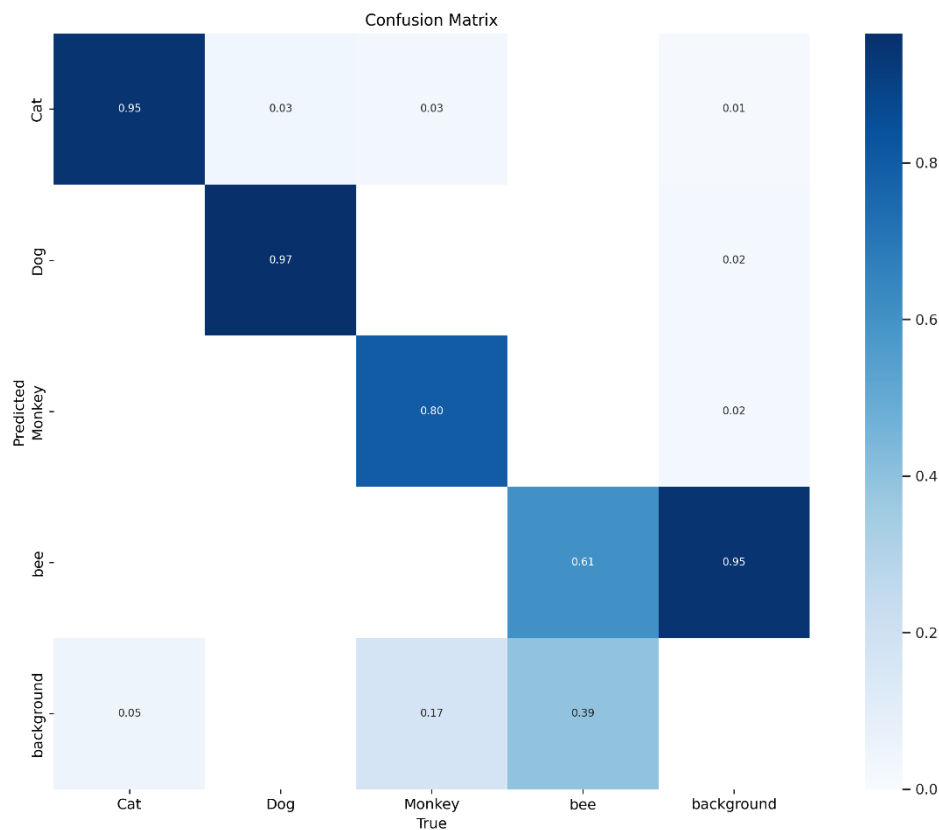


Gambar 4. Grafik Loss dan Metriks Training dan Validasi Model

Plot train/box\_loss menunjukkan penurunan loss untuk regresi bounding box dari nilai 0,08 hingga 0,04 pada akhir pelatihan. Sementara pada plot train/obj\_loss, nilai loss untuk deteksi objek menurun dari sekitar 0,0325 hingga 0,01 di akhir proses. Pada metrik klasifikasi yakni train/cls\_loss, nilainya menurun dari 0,04 hingga mendekati 0,01 di akhir pelatihan. Di sisi lain, metrik seperti precision dan recall untuk evaluasi performa model meningkat dari nilai rendah di awal hingga mendekati 0,8 di akhir proses. Begitu pula metrik mAP (mean Average Precision) untuk deteksi objek meningkat dari nilai sekitar 0,3 hingga mendekati 0,8 pada akhir pelatihan. Penurunan loss dan peningkatan metrik seperti precision, recall, dan mAP mengindikasikan peningkatan performa model dalam melakukan deteksi dan klasifikasi objek seiring berjalannya proses pelatihan.

### 3.3. Confusion Matrix

Confusion matrix atau matriks konfusi yang menunjukkan hasil klasifikasi dari sebuah model machine learning untuk mengklasifikasikan beberapa objek, yaitu Cat, Dog, Monkey, Bee, dan Background. Matriks konfusi menyajikan informasi tentang klasifikasi aktual dan prediksi yang dilakukan oleh model dalam bentuk matriks. Diagonal utama matriks menunjukkan jumlah sampel yang diklasifikasikan dengan benar untuk setiap kelas, sedangkan nilai-nilai di luar diagonal utama merupakan kesalahan klasifikasi. Berdasarkan gambar, dapat dilihat bahwa model memiliki akurasi tertinggi dalam mengklasifikasikan kelas Cat dengan nilai 0,95 atau 95%. Sementara untuk kelas Dog dan Monkey, akurasi model masing-masing adalah 0,97 atau 97% dan 0,80 atau 80%. Namun, model tampaknya kurang akurat dalam mengklasifikasikan kelas Bee, dengan nilai 0,61 atau 61% sampel Bee diklasifikasikan dengan benar, sementara 0,39 atau 39% lainnya salah diklasifikasikan. Selain itu, terdapat beberapa kesalahan klasifikasi di luar diagonal utama, seperti 0,03 sampel Cat salah diklasifikasikan sebagai Dog dan Monkey, serta 0,02 sampel Dog dan Monkey salah diklasifikasikan sebagai Bee.



Gambar 5. Confusion Matrix

#### 4. Kesimpulan

Penelitian ini berfokus pada implementasi model You Only Look Once (YOLO) untuk deteksi objek, khususnya hewan peliharaan. Model YOLO, yang berbasis deep learning, menawarkan beberapa keunggulan dibandingkan metode tradisional, termasuk prediksi bounding box dan probabilitas kelas objek secara simultan, arsitektur Convolutional Neural Network (CNN) yang relatif sederhana, dan kecepatan komputasi yang tinggi, sehingga cocok untuk aplikasi real-time. Studi ini menggunakan dataset sebanyak 770 gambar, dengan 524 untuk pelatihan, 136 untuk validasi, dan 110 untuk pengujian. Data gambar melalui proses anotasi untuk memberi label pada daerah yang menjadi perhatian dengan label kelas. Proses pelatihan melibatkan optimalisasi model YOLO menggunakan data yang dianotasi, sementara metrik kinerja seperti presisi, recall, dan mean Average Precision (mAP) diperoleh melalui proses validasi. Hasil menunjukkan kemampuan model dalam mendeteksi dan mengklasifikasikan objek secara efektif, dengan metrik kinerja seperti presisi, recall, dan mAP mendekati 0,8 di akhir proses pelatihan. Fungsi kerugian untuk regresi bounding box, deteksi objek, dan klasifikasi juga menunjukkan tren penurunan, mengindikasikan peningkatan kinerja model. Selanjutnya, sebuah confusion matrix disajikan, memberikan wawasan tentang akurasi model dalam mengklasifikasikan kelas yang berbeda. Model mencapai akurasi tertinggi untuk kelas 'Cat' sebesar 95%, sementara kelas 'Bee' menjadi yang paling menantang dengan akurasi 61%. Secara keseluruhan, implementasi model YOLO dalam studi ini menunjukkan efektivitasnya dalam mendeteksi dan mengklasifikasikan hewan peliharaan. Namun, masih ada ruang untuk peningkatan, terutama dalam meningkatkan akurasi model untuk kelas yang menantang dan mengatasi kemungkinan kesalahan klasifikasi.

### Daftar Pustaka

- [1] R. Dwiyanto, D. W. Widodo, and P. Kasih, "Implementasi Metode You Only Look Once (YOLOv5) Untuk Klasifikasi Kendaraan Pada CCTV Kabupaten Tulungagung," in Seminar Nasional Inovasi Teknologi, 2022, pp. 102-104.
- [2] A. R. Wasril, M. S. Ghozali, and M. B. Mustafa, "Pembuatan Pendeteksi Obyek Dengan Metode You Only Look Once (YOLO) Untuk Automated Teller Machine (ATM)," Majalah Ilmiah UNIKOM, vol. 17, no. 1, pp. 69-75, 2019.
- [3] Janapriya, A.A.G.B., 2023. Pengenalan Jenis Rambu Lalu Lintas menggunakan Metode YOLO V5. Jurnal Elektronik Ilmu Komputer Udayana, 11(4), pp.1011-1018.
- [4] Priana, K.A., Karyawati, A.A.I.N.E., 2023. Sistem Pendeteksi Sampah Secara Realtime Menggunakan Metode YOLO. Jurnal Nasional Teknologi Informasi dan Aplikasinya, 2(1), pp.31-36.