



Penerapan Metode YOLOv5 untuk Klasifikasi dan Deteksi Objek Menggunakan Video Non-Real-Time

Yovi Apridiansyah¹, Zeko Padli², Yuza Reswan³, Harsry Witriyono⁴

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu, Jl. Bali, Kampung Bali, Teluk Segara, Kota Bengkulu, 38119, Indonesia .

Email: yovi.apridiansyah@umb.ac.id¹, zekobkl09@gmail.com², yuzareswan@umb.ac.id³, harrywitriyono@umb.ac.id⁴

Abstrak– Deteksi objek merupakan salah satu penerapan utama dari teknologi *computer vision* dalam bidang kecerdasan buatan. Salah satu algoritma deteksi objek yang banyak digunakan karena efisiensinya adalah YOLOv5. Penelitian ini bertujuan untuk menerapkan metode YOLOv5 dalam mendeteksi dan mengklasifikasikan objek kendaraan dan manusia pada rekaman video non-real-time di kawasan Simpang Tugu Pena Kota Bengkulu. Dataset yang digunakan merupakan video hasil dokumentasi lapangan, yang kemudian dianalisis menggunakan model YOLOv5 dengan pelatihan berbasis transfer learning. Untuk menjaga identitas objek antar-frame, sistem dikombinasikan dengan metode Kalman Filter dan SORT sebagai pelacak objek. Hasil pengujian menunjukkan bahwa model yang dibangun mampu melakukan deteksi objek dengan baik pada kondisi visual yang bervariasi, serta mencapai nilai akurasi yang tinggi berdasarkan pengukuran menggunakan matriks konfusi. Penelitian ini menunjukkan bahwa penerapan YOLOv5 efektif digunakan dalam sistem dokumentasi visual berbasis AI di lingkungan ruang publik yang dinamis.

Kata kunci: YOLOv5; deteksi objek; Kalman Filter; SORT; Simpang Tugu Pena; computer vision.

Abstract– Object detection is one of the main applications of computer vision technology in the field of artificial intelligence. One object detection algorithm widely used due to its efficiency is YOLOv5. This study aims to apply the YOLOv5 method to detect and classify vehicle and human objects in non-real-time video recordings in the Tugu Pena Intersection area of Bengkulu City. The dataset used is field documentation videos, which are then analyzed using the YOLOv5 model with transfer learning-based training. To maintain object identity between frames, the system is combined with the Kalman Filter and SORT methods as object trackers. Test results show that the developed model is capable of good object detection under varying visual conditions and achieves high accuracy based on measurements using a confusion matrix. This study demonstrates the effectiveness of YOLOv5 in an AI-based visual documentation system in dynamic public space environments.

Keywords: YOLOv5; object detection; Kalman Filter; SORT; Tugu Pena Intersection; computer vision.

1. PENDAHULUAN

Tugu Pena adalah ikon bersejarah sekaligus simpang lalu lintas utama di Kota Bengkulu. Kawasan ini ramai oleh kendaraan dan aktivitas masyarakat setiap hari. Kondisi tersebut membuat lalu lintas di sekitar Tugu Pena padat dan dinamis. Situasi ini berpotensi menimbulkan kemacetan dan kecelakaan. Oleh karena itu, dibutuhkan sistem pemantauan visual yang cerdas, otomatis, dan efisien. Sistem ini bermanfaat untuk dokumentasi, perencanaan lalu lintas, serta keamanan. Teknologi computer vision dan kecerdasan buatan (AI) menjadi solusi potensial. Metode deteksi objek seperti YOLOv5 mampu mengidentifikasi kendaraan dan manusia secara cepat dan akurat pada rekaman video non-real-time.

Salah satu algoritma deteksi objek paling populer adalah YOLO (*You Only Look Once*). YOLOv4 dikenal karena menyeimbangkan kecepatan dan akurasi dalam pemrosesan data visual [1]. Versi selanjutnya, YOLOv5, hadir dengan peningkatan efisiensi dan arsitektur yang lebih ringan. Menurut Alhassan dan Yilmaz, YOLOv5 terbukti lebih unggul dari YOLOv4 dalam hal presisi dan kecepatan pada sistem pengawasan berbasis UAV [2].

Dalam konteks lalu lintas padat, Rahman dkk. menggunakan YOLOv5 untuk mendeteksi kendaraan di area perkotaan dan memperoleh akurasi tinggi meskipun dalam kondisi visual kompleks [3]. Di bidang industri, Simeth dkk. menerapkan YOLOv5 dalam lini produksi manufaktur dan mendapatkan hasil deteksi yang konsisten dan andal [4]. Selain itu, Rahayu membuktikan efektivitas YOLOv5 dalam klasifikasi ras kucing berdasarkan metadata dari gambar dataset [5].

Agar sistem deteksi antar-frame lebih stabil, algoritma pelacakan seperti Kalman Filter dan SORT digunakan. Kalman Filter berguna untuk mencerminkan posisi objek pada frame berikutnya secara matematis [6], sedangkan SORT memungkinkan pelacakan multi-objek secara efisien dan real-time menggunakan prediksi

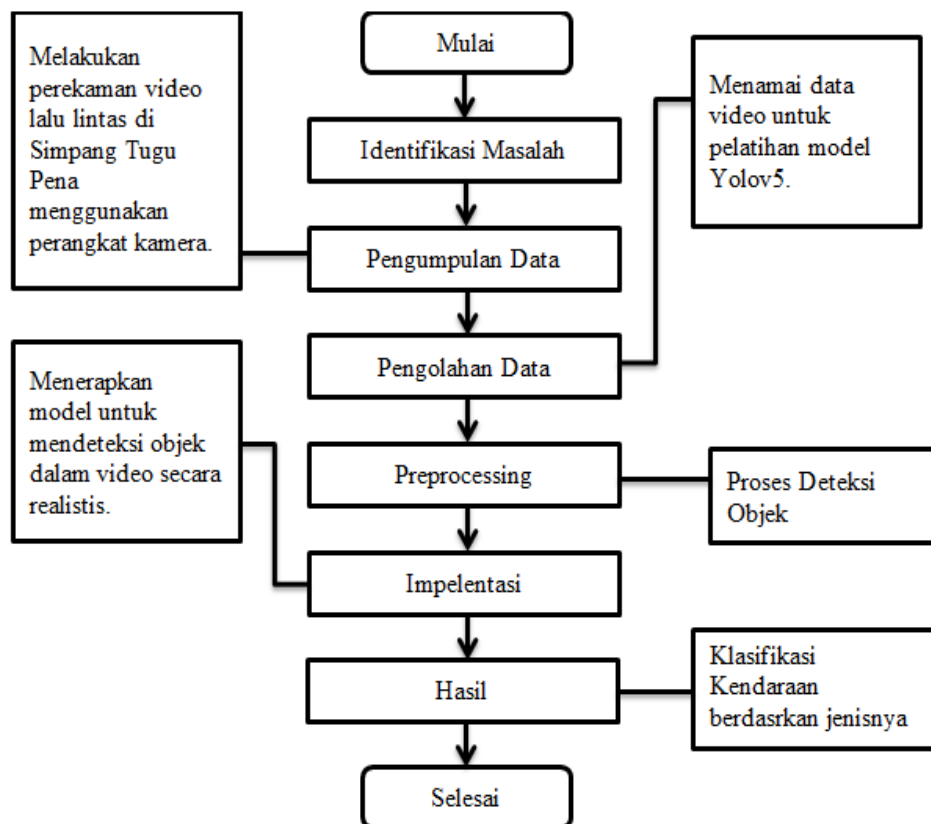
Kalman dan metode asosiasi [7]. Lu dkk. meningkatkan performa YOLOv5 dengan mengintegrasikan Swin-Transformer dalam mendeteksi buah anggur, yang terbukti meningkatkan akurasi terutama dalam pencahayaan bervariasi [8]. Sementara itu, Mulyana dkk. menggunakan YOLOv5 dalam pendeteksian bahasa isyarat huruf hijaiyah, menunjukkan model yang pendek dalam mengenali pola gerakan tangan [9].

Dalam studi berbasis UAV, Zhang et al. menerapkan YOLOv5 untuk mendeteksi objek dari citra udara dan memperoleh hasil yang memuaskan dalam klasifikasi objek kecil di ketinggian [10]. Untuk penerapan protokol kesehatan, Ashar dan Suarna memanfaatkan YOLOv5 dalam mendeteksi penggunaan masker di lingkungan kantor sebagai bentuk kepatuhan secara otomatis [11]. Sebagai penguatan teoritis, Terven dkk. menyusun ulasan komprehensif tentang seluruh evolusi arsitektur YOLO, dari YOLOv1 hingga YOLOv8, dan menyebut YOLOv5 sebagai model yang sangat seimbang antara efisiensi dan kecepatan [12].

Melalui beragam penelitian tersebut, YOLOv5 terbukti fleksibel, ringan, dan akurat untuk berbagai skenario deteksi objek, baik dalam konteks transportasi, industri, citra udara, maupun kebutuhan sosial [13], [14], [15]. Oleh karena itu, penelitian ini bertujuan menerapkan YOLOv5 dalam mendeteksi dan mengklasifikasikan objek di kawasan Simpang Tugu Pena Kota Bengkulu, serta mengombinasikannya dengan metode Kalman Filter dan SORT untuk menjaga konsistensi identitas objek antar-frame dalam video non-real-time.

2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini disusun berdasarkan pertimbangan efektivitas dalam pemrosesan visual berbasis AI. Framework PyTorch dipilih karena memiliki dokumentasi luas, mendukung GPU acceleration, serta kompatibel dengan model YOLOv5 yang dikembangkan oleh Ultralytics. Penggunaan Kalman Filter dan SORT dalam pelacakan objek bertujuan untuk mempertahankan identitas objek secara konsisten antar-frame, mengingat video yang digunakan bersifat non-real-time dan memiliki potensi gerakan objek yang kompleks. Alur proses penelitian ini ditunjukkan pada diagram kerangka kerja yang menggambarkan tahapan utama mulai dari pengumpulan data hingga evaluasi sistem.



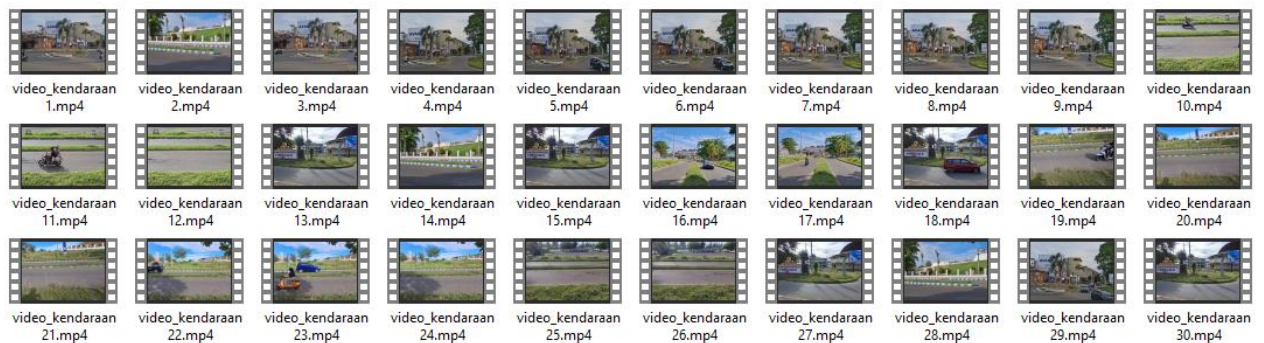
Gambar 1. Diagram Kerangka Kerja Penelitian

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen dengan pengujian sistem deteksi objek berbasis video statis. Penelitian dilakukan dalam beberapa tahap mulai dari pengumpulan data, pemrosesan data, pelatihan model, hingga evaluasi hasil.

2.2 Pengumpulan Data

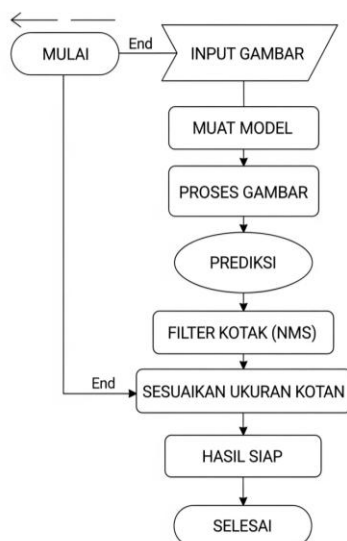
Data uji dalam penelitian ini berupa 30 video yang direkam di kawasan Simpang Tugu Pena. Video tersebut memiliki durasi rata-rata 10 detik dengan variasi kondisi pencahayaan (siang dan sore) serta tingkat keramaian lalu lintas. Setiap video kemudian diekstrak menjadi frame untuk dijadikan dataset anotasi. Data ini berfungsi sebagai input utama dalam proses pelatihan dan pengujian model YOLOv5. Jumlah serta variasi video uji ditampilkan pada Gambar 2.2.



Gambar 2. 30 Video Data Uji

2.3 Pra-Pemrosesan Data

Pra-pemrosesan data bertujuan untuk menyiapkan input yang sesuai dengan kebutuhan arsitektur YOLOv5. Tahapan ini mencakup transformasi gambar dari video menjadi bentuk data yang siap dilatih oleh model. Proses ini penting untuk memastikan bahwa kualitas data tetap terjaga dan model dapat belajar secara optimal dari setiap frame yang dianotasi. Alur tahapan pra-pemrosesan tersebut dapat dilihat pada Gambar 3 yang menampilkan proses mulai dari pemuatan data, konversi, normalisasi, hingga data siap digunakan untuk pelatihan model.



Gambar 3. Diagram tahapan pra-pemrosesan data

Frame yang diperoleh dari video diproses melalui beberapa tahapan: konversi format gambar ke tensor PyTorch, normalisasi nilai piksel, resizing ke dimensi input YOLOv5 (640x640), serta augmentasi data (rotasi, flipping, penyesuaian brightness dan kontras). Frame juga diberi anotasi manual menggunakan tool labeling untuk membentuk ground truth.

2.4 Pelatihan Model

Model dibuat menggunakan framework PyTorch dengan modul pelatihan dari Ultralytics YOLOv5. Dataset dibagi menjadi dua bagian, yaitu data latih (data pelatihan) dan validasi data (validation data). Data latih digunakan agar model dapat mempelajari pola objek, sedangkan validasi data digunakan untuk menguji kinerja model pada data yang belum pernah dilihat sehingga menghindari overfitting [16].

Selama proses pelatihan, beberapa metrik evaluasi digunakan. Presisi mengukur tepat prediksi positif, yaitu perbandingan antara jumlah deteksi benar dengan seluruh deteksi positif yang dihasilkan. Ingat model ukuran kemampuan menemukan semua objek yang relevan. Kedua metrik umum digunakan dalam menilai kinerja sistem klasifikasi [17]. Selain itu, akurasi juga digunakan untuk mengukur proporsi prediksi benar terhadap keseluruhan data, sehingga menggambarkan kinerja model secara umum [18].

Proses pelatihan dilakukan secara iteratif (epoch) hingga model mencapai stabilitas pada nilai presisi , recall , dan akurasi yang optimal.

2.5 Pelacakan dan Evaluasi

Untuk pelacakan objek, digunakan algoritma Kalman Filter dan SORT agar sistem dapat menjaga konsistensi identitas objek antar frame. Visualisasi hasil dilakukan dengan menampilkan bounding box dan label pada tiap objek. Evaluasi sistem dilakukan menggunakan confusion matrix serta menghitung precision, recall, dan akurasi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi

Confusion matrix atau matriks kebingungan adalah metode evaluasi yang digunakan untuk menilai performa model klasifikasi dengan cara membandingkan hasil prediksi model terhadap data aktual. Matriks ini berbentuk tabel dua dimensi yang memuat informasi jumlah prediksi yang benar maupun salah pada setiap kelas [19].

Dalam confusion matrix terdapat empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Keempat komponen tersebut menjadi dasar dalam menghitung ukuran performa model. Beberapa ukuran yang umum digunakan adalah akurasi, presisi, dan recall [20].

1. Akurasi

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Akurasi mengukur proporsi data yang diprediksi dengan benar dibandingkan dengan keseluruhan data uji.

2. Presisi

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

Presisi menunjukkan tingkat ketepatan model dalam memprediksi data positif.

3. Recall (Sensitivitas)

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Recall menggambarkan kemampuan model dalam menemukan seluruh data positif yang sebenarnya ada.

	POSITIF	NEGATIF
POSITIF	TP = 17	FP = 3
NEGATIF	FN = 5	TN = 5

Gambar 4. Confusion Matrix Model YOLOv5

Pada Gambar 4 terlihat bahwa diagonal utama matriks mewakili jumlah data yang diklasifikasikan dengan benar (TP dan TN), sementara bagian di luar diagonal menunjukkan kesalahan klasifikasi (FP dan FN). Dengan demikian, confusion matrix dapat memberikan gambaran detail mengenai distribusi hasil prediksi.

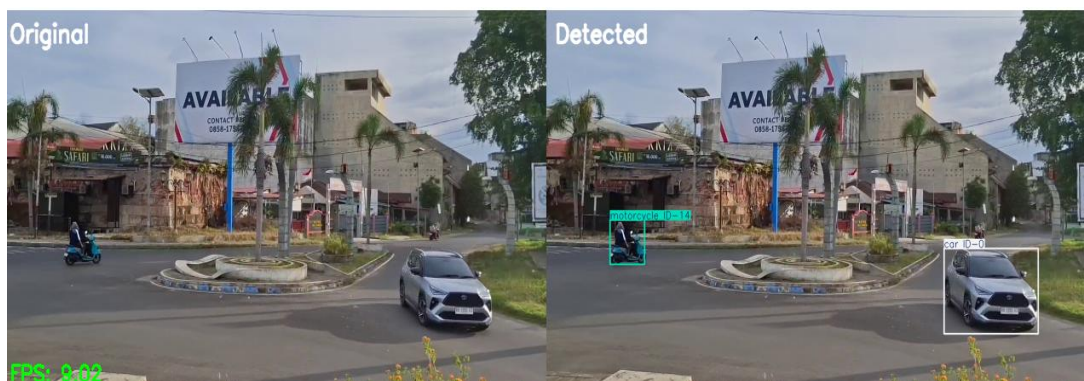
Tabel 3.1 Hasil Perhitungan Confusion Matrix

Metrik	Nilai
Akurasi	73,3 %
Presisi	85 %
Recall	77,3 %

Berdasarkan hasil pada Tabel 3.1, diperoleh **akurasi sebesar 73,3%**, yang berarti model mampu mengklasifikasikan data dengan benar pada 73,3% dari total data uji. Nilai **presisi sebesar 85%** menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan model sesuai dengan kondisi aktual, sehingga tingkat kesalahan prediksi positif (False Positive) relatif rendah. Selanjutnya, nilai **recall sebesar 77,3%** menandakan bahwa model cukup baik dalam menemukan data positif yang sebenarnya ada, meskipun masih terdapat beberapa data positif yang terlewat (False Negative).

Secara keseluruhan, kombinasi nilai akurasi, presisi, dan recall tersebut menunjukkan bahwa model memiliki performa yang cukup baik, dengan kecenderungan lebih kuat pada ketepatan prediksi (presisi) dibandingkan kelengkapan deteksi (recall). Evaluasi menggunakan confusion matrix membantu memberikan gambaran yang lebih rinci dibandingkan hanya mengandalkan akurasi semata [21].

3.2 Analisis Sistem



Gambar 5. Contoh Hasil Deteksi Objek

Berdasarkan Gambar 5 bahwa Sistem mampu mendeteksi dan melacak objek secara simultan. Kombinasi antara YOLOv5 dan SORT memungkinkan identifikasi pergerakan objek dan klasifikasi kelas secara konsisten, sedangkan Kalman Filter digunakan untuk prediksi posisi objek antar frame. dan klasifikasi kelas secara konsisten. Penerapan Non-Maximum Suppression (NMS) membantu menyaring prediksi yang tumpang tindih.

Model menunjukkan ketahanan terhadap variasi pencahayaan dan latar belakang dinamis. Namun, masih ditemukan kesalahan deteksi pada beberapa objek kecil atau terhalang. Peningkatan kualitas dataset dan augmentasi lanjutan dapat mengatasi kendala tersebut.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan metode YOLOv5 pada video non-real-time di kawasan Simpang Tugu Pena Kota Bengkulu mampu melakukan deteksi dan klasifikasi objek kendaraan serta manusia dengan performa yang cukup baik, ditunjukkan oleh nilai akurasi sebesar 73,3%, presisi 85%, dan recall 77,3%. Integrasi metode pelacakan Kalman Filter dan SORT membantu menjaga konsistensi identitas objek antar-frame, sehingga sistem mampu melacak pergerakan secara simultan meskipun kondisi visual dinamis. Hasil ini membuktikan bahwa YOLOv5 efektif digunakan sebagai sistem pemantauan visual berbasis AI di ruang publik, meskipun masih terdapat keterbatasan dalam mendeteksi objek kecil atau tertutup sebagian, yang dapat diatasi melalui peningkatan kualitas dataset dan pengembangan model lebih lanjut untuk aplikasi real-time maupun integrasi dengan perangkat pemantauan seperti CCTV dan drone.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Bengkulu dan dosen pembimbing atas dukungan selama penelitian ini berlangsung.

REFERENCES

- [1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020.
- [2] M. A. M. Alhassan and E. Yilmaz, "Evaluating YOLOv4 and YOLOv5 for Enhanced Object Detection in UAV-Based Surveillance," *Processes*, vol. 13, no. 1, 2025, doi: 10.3390/pr13010254.
- [3] R. Rahman, Z. Bin Azad, and M. Bakhtiar Hasan, "Densely-Populated Traffic Detection Using YOLOv5 and Non-maximum Suppression Ensembling," *Lect. Notes Data Eng. Commun. Technol.*, vol. 95, pp. 567–578, 2022, doi: 10.1007/978-981-16-6636-0_43.
- [4] A. Simeth, A. A. Kumar, and P. Plapper, "Flexible and robust detection for assembly automation with YOLOv5: a case study on HMLV manufacturing line," *J. Intell. Manuf.*, vol. 36, no. 5, pp. 3447–3463, 2024, doi: 10.1007/s10845-024-02411-5.
- [5] M. Ismu Rahayu, "KLASIFIKASI RAS KUCING MENGGUNAKAN METADATA DATASET KAGGLE DENGAN FRAMEWORK YOLO v5," *J. Teknol. Inf. dan Komun.*, vol. 12, no. 1, pp. 14–18, 2023, doi: 10.58761/jurtikstmikbandung.v12i1.179.
- [6] M. Soeleman, D. Raniasti, K. W. M. Muljono, M. Muslih, and R. Pramunendar, "Tracking Moving Objects based on Background Subtraction using Kalman Filter," pp. 1–8, 2021, doi: 10.4108/eai.24-10-2018.2280505.
- [7] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2020–August, pp. 3464–3468, 2020, doi: 10.1109/ICIP.2016.7533003.
- [8] S. Lu, X. Liu, Z. He, X. Zhang, W. Liu, and M. Karkee, "Swin-Transformer-YOLOv5 for Real-Time Wine Grape Bunch Detection," *Remote Sens.*, vol. 14, no. 22, pp. 0–2, 2022, doi: 10.3390/rs14225853.
- [9] D. I. Mulyana, M. F. Lazuardi, and M. B. Yel, "Deteksi Bahasa Isyarat Dalam Pengenalan Huruf Hijaiyah Dengan Metode YOLOv5," *J. Tek. Elektro dan Komputasi*, vol. 4, no. 2, pp. 145–151, 2022.
- [10] H. Zhang, F. Shao, X. He, Z. Zhang, Y. Cai, and S. Bi, "Research on Object Detection and Recognition Method for UAV Aerial Images Based on Improved YOLOv5," *Drones*, vol. 7, no. 6, 2023, doi: 10.3390/drones7060402.
- [11] M. H. Ashar and D. Suarna, "KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Algoritma YOLOv5 dalam Mendeteksi Penggunaan Masker Pada Kantor Biro Umum Gubernur Sulawesi Barat," *Media Online*, vol. 3, no. 3, pp. 298–302, 2022.
- [12] J. Terven, D. M. Córdova-Esparza, and J. A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.
- [13] Yashaswini and Dr. Manimala, "Classification and Detections using Yolov5," *Int. J. Multidiscip. Res.*, vol. 5, no. 5, pp. 1–3, 2023, doi: 10.36948/ijfmr.2023.v05i05.6057.
- [14] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021–Octob, pp. 2778–2788, 2021, doi: 10.1109/ICCVW54120.2021.00312.
- [15] M. Hussain, "YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-time Vision," pp. 1–12, 2024.

- [16] Z. Gao and X. Wang, "Deep learning," *EEG Signal Process. Featur. Extr.*, pp. 325–333, 2019, doi: 10.1007/978-981-13-9113-2_16.
- [17] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," no. January 2007, 2020, [Online]. Available: <http://arxiv.org/abs/2010.16061>
- [18] C. Robert, "Machine Learning, a Probabilistic Perspective," *Chance*, vol. 27, no. 2, pp. 62–63, 2014, doi: 10.1080/09332480.2014.914768.
- [19] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African J. Biomed. Res.*, vol. 27, no. 4, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [20] "005ConfusionMatrixandPerformanceEvaluationMetrics."
- [21] A. Farhan AlShammari, "Implementation of Model Evaluation using Confusion Matrix in Python," *Int. J. Comput. Appl.*, vol. 186, no. 50, pp. 975–8887, 2024.