JTERA (Jurnal Teknologi Rekayasa), Vol. 10, No. 1, Juni 2025, Hal. 105-112 p-ISSN 2548-737X
Terakreditasi "**Peringkat 3**" oleh Kemdiktisaintek, Nomor : 10/C/C3/DT.05.00/2025 e-ISSN 2548-8678
DOI: 10.31544/jtera.v10.i1.2025.105-112

Metode Anotasi Visual Baru untuk Meningkatkan Akurasi Sistem Penentuan Posisi dalam Ruangan Berbasis Kamera

Lathifa Nur Ramdhania[1], Hanif Almadaniy[2], Ahmad Aminudin[3], Lilik Hasanah[4]
[1,2,3,4] Program Studi Fisika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia
Jl. Dr. Setiabudi No.229, Isola, Kec. Sukasari, Kota Bandung, Jawa Barat 40154, Indonesia
lathifa@upi.edu

Abstrak

Sistem penentuan posisi dalam ruangan (*Indoor Positioning System*/IPS) memiliki peranan penting dalam berbagai aplikasi seperti navigasi robot, pelacakan aset, dan sistem layanan otomatis. Salah satu pendekatan IPS yang menjanjikan adalah berbasis kamera menggunakan algoritma YOLOv4 karena biayanya yang relatif rendah dan fleksibilitas penggunaannya, namun akurasinya masih tergolong rendah. Studi ini meningkatkan akurasi sistem IPS berbasis kamera melalui pendekatan anotasi visual baru, yaitu dengan menempatkan anotasi pada titik tengah antara dua kaki manusia sebagai acuan posisi objek di lantai. Dataset khusus berisi 2.000 citra digunakan untuk melatih model YOLOv4 yang telah disesuaikan. Hasil evaluasi menunjukkan peningkatan performa dengan nilai mean average precision (mAP) sebesar 99,19% setelah pelatihan sebanyak 6.000 iterasi. Rasio konversi pixel-ke-centimeter yang diperoleh mencapai 0,309 cm/pixel (sumbu-x) dan 0,308 cm/pixel (sumbu-y), dengan peningkatan akurasi sebesar 62,43% dan penurunan standar deviasi sebesar 69,01%. Temuan ini menunjukkan bahwa metode kalibrasi menggunakan anotasi pada kaki secara signifikan meningkatkan akurasi estimasi posisi dalam ruangan.

Kata kunci: *indoor positioning system, computer vision,* deteksi objek berbasis kamera, algoritma YOLO, kalibrasi visual,

Abstract

Indoor Positioning Systems (IPS) play a crucial role in various applications such as robot navigation, asset tracking, and automated service systems. One promising IPS approach is camera-based positioning using the YOLOv4 algorithm due to its relatively low cost and flexible implementation; however, its accuracy remains relatively low. This study enhances the accuracy of camera-based IPS by introducing a new visual annotation approach, which uses the midpoint between a person's feet as the reference point for object positioning on the floor. A dedicated dataset consisting of 2,000 images was used to train a customized YOLOv4 model. Evaluation results show improved performance with a mean average precision (mAP) of 99.19% after 6,000 training iterations. The resulting pixel-to-centimeter conversion ratios were 0.309 cm/pixel (x-axis) and 0.308 cm/pixel (y-axis), with an accuracy improvement of 62.43% and a standard deviation reduction of 69.01%. These findings demonstrate that calibration using foot midpoint annotations significantly enhances the accuracy of indoor position estimation.

Keywords: *indoor positioning system, computer vision, camera-based object detection, YOLO algorithm, object annotation,*

I. PENDAHULUAN

Sistem deteksi objek dalam ruangan (*Indoor Positioning System* atau IPS) merupakan teknologi yang dirancang untuk menentukan lokasi objek di dalam suatu bangunan secara real-time, terutama pada lingkungan yang tidak terjangkau oleh sinyal *Global Positioning System* (GPS) [1]. Akurasi GPS menurun drastis saat digunakan di dalam ruangan karena sinyal satelit terhalang oleh struktur bangunan. Untuk mengatasi keterbatasan tersebut, IPS dengan berbagai pendekatan telah dikembangkan, antara lain IPS berbasis *ultrawideband* (UWB), Wi-Fi, sensor ultrasonik, dan kamera [1]. Di antara pendekatan-pendekatan ini, sistem berbasis kamera dinilai menjanjikan karena

tidak memerlukan infrastruktur tambahan dan memiliki biaya implementasi yang relatif rendah, meskipun menuntut pemrosesan visual yang kompleks [2]. IPS saat ini menjadi teknologi yang banyak digunakan dalam berbagai aplikasi seperti navigasi robot [3], [4], [5], Augmented Reality (AR) [6], [7], serta pelacakan aset di gudang dan area industri [8], [9].

Dalam dekade terakhir, kemajuan pesat dalam bidang *computer vision* dan *deep learning* telah mendorong penerapan metode *convolutional neural network* (CNN) untuk berbagai kebutuhan persepsi visual, termasuk dalam sistem IPS berbasis kamera[10], [11]. Algoritma deteksi objek seperti *You Only Look Once* (YOLO) telah terbukti efektif dalam mendeteksi objek secara *real-time* dengan kecepatan tinggi dan akurasi yang kompetitif [12], [13].

YOLO merupakan salah satu algoritma deteksi objek berbasis CNN yang mengintegrasikan seluruh proses deteksi, mulai dari ekstraksi fitur hingga prediksi kelas dan posisi objek ke dalam satu langkah terpadu[12]. Berbeda dengan metode konvensional seperti *sliding window* atau *region proposal*, YOLO melakukan deteksi secara menyeluruh terhadap citra dalam satu kali pemrosesan, sehingga lebih cepat dan efisien.

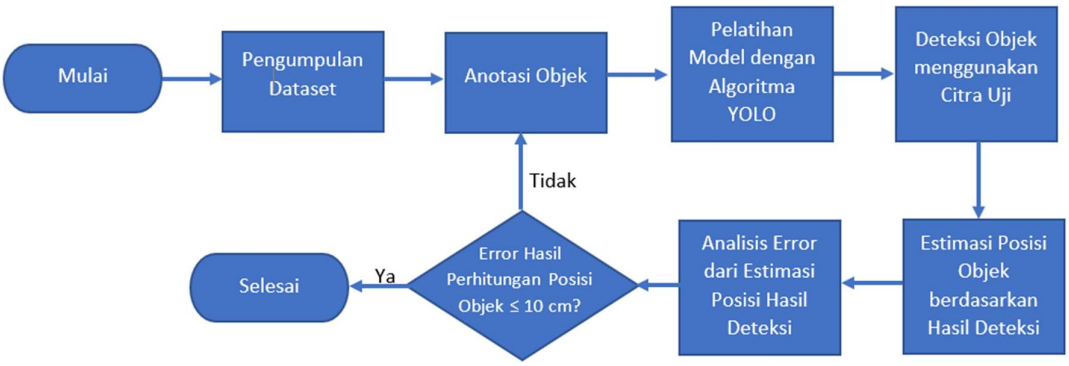
Berbagai studi telah mengeksplorasi pemanfaatan varian YOLO, termasuk YOLOv4 dan YOLOv5, untuk mendeteksi posisi manusia atau objek di dalam ruangan [14], [15], [16]. Salah satu studi oleh Kurniawan [17] mengembangkan sistem IPS berbasis kamera *smartphone* menggunakan model YOLOv4. Meskipun berhasil mendeteksi keberadaan manusia, sistem tersebut menghasilkan *error* posisi rata-rata sebesar 40,14 cm dari target 10 cm, sehingga belum memenuhi kebutuhan presisi pada aplikasi seperti pengambilan barang otomatis atau navigasi di ruang sempit.

Salah satu kelemahan dari pendekatan sebelumnya adalah belum optimalnya metode kalibrasi visual, di mana deteksi objek belum secara langsung dikonversi ke koordinat fisik dengan akurasi tinggi. Deteksi *bounding box* objek, seperti tubuh manusia, tidak serta-merta menunjukkan posisi aktual objek di lantai, terutama pada tampilan citra dari atas. Oleh karena itu, diperlukan pendekatan anotasi yang lebih tepat untuk menjadikan deteksi objek sebagai dasar konversi pixel objek ke satuan panjang.

Untuk mengatasi keterbatasan akurasi pada pendekatan sebelumnya, studi ini mengusulkan sistem IPS berbasis kamera yang diintegrasikan dengan metode kalibrasi visual, yaitu dengan memanfaatkan anotasi pada titik tengah antara dua kaki manusia. Titik ini dipilih karena secara visual lebih stabil dan secara geometris merepresentasikan proyeksi posisi manusia di lantai dalam citra tampak atas. Dataset khusus yang terdiri dari 2.000 citra digunakan untuk melatih model YOLOv4 yang telah dimodifikasi, dan dievaluasi menggunakan rasio konversi pixel ke centimeter. Pendekatan ini diharapkan mampu menurunkan *error* posisi secara signifikan dan menghasilkan sistem IPS yang lebih akurat serta layak diterapkan pada aplikasi-aplikasi yang membutuhkan presisi tinggi.

II. METODE PENELITIAN

Penelitian ini terdiri atas lima tahapan utama sebagaimana ditunjukkan pada Gambar 1, yaitu pengumpulan dataset, anotasi objek, pelatihan model menggunakan algoritma YOLOv4, deteksi posisi objek, estimasi posisi objek berdasarkan hasil deteksi dan evaluasi akurasi hasil deteksi.



Gambar 1. Diagram Alir Penelitian

2.1. Pengumpulan Data Citra

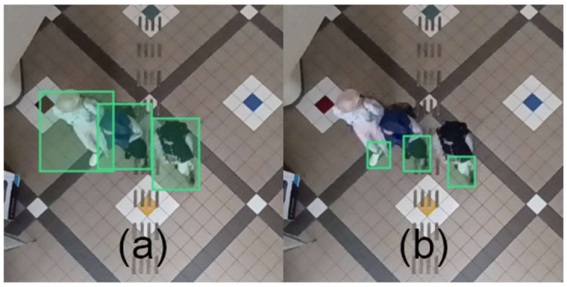
Data citra dalam penelitian ini didefinisikan sebagai file gambar berformat JPEG atau JPG yang menampilkan citra manusia dari sudut pandang atas (*top view*). Dataset terdiri dari file citra yang dipasangkan dengan file teks yang berisi informasi anotasi dalam format YOLO. Sumber data citra berasal dari dua kategori utama: (1) dataset dari penelitian sebelumnya yang relevan diperoleh menggunakan kamera *smartphone* dengan posisi

kamera dipasang secara tetap, dan (2) hasil pencarian daring menggunakan kata kunci seperti *"overhead"*, *"human overhead"*, dan *"top view"*, dengan penyaringan untuk mengecualikan citra yang mengandung efek *fisheye*. Posisi kamera yang tidak berubah selama pengambilan citra membuat distorsi perspektif dianggap minimal dan konstan, sehingga kalibrasi visual berdasarkan dimensi ruangan dapat digunakan secara konsisten. Dataset hasil seleksi berjumlah 2.000 citra, yang seluruhnya dianotasi ulang untuk mempersiapkan data pelatihan model.

2.2. Anotasi Objek

Setiap citra dalam dataset perlu dianotasi terlebih dahulu, yaitu diberi informasi mengenai lokasi objek dalam bentuk *bounding box* yang mengelilingi kedua kaki manusia. Anotasi ini berfungsi sebagai referensi bagi model untuk mempelajari pola visual dan posisi objek dalam gambar. Format anotasi yang digunakan mengikuti standar YOLO, di mana setiap gambar memiliki pasangan file teks yang berisi label objek dan koordinat *bounding box* dalam satuan relatif terhadap dimensi gambar.

Proses anotasi dilakukan secara manual menggunakan perangkat lunak *makesense.ai* dengan menandai kedua kaki manusia dalam setiap citra menggunakan *bounding box*. Titik tengah pada setiap *bounding box* ditentukan dan digunakan sebagai representasi posisi objek pada lantai. Penyesuaian anotasi titik tengah antar kaki manusia sebagai *anchor point* ini dilakukan sehingga model tidak hanya mendeteksi keberadaan objek, tetapi juga mampu memprediksi posisi relatif objek secara lebih akurat dalam sistem koordinat gambar. Gambar 2 memperlihatkan perbandingan antara metode anotasi dari studi terdahulu dan pendekatan anotasi baru yang diterapkan dalam penelitian ini.



Gambar 2. Perbandingan metode anotasi: (a) anotasi berdasarkan *bounding box* keseluruhan tubuh (studi sebelumnya); (b) anotasi dengan titik tengah kaki sebagai referensi posisi (pendekatan studi ini).

2.3. Pelatihan Model

Model deteksi objek dilatih menggunakan 2.000 citra yang telah dianotasi menggunakan algoritma YOLOv4. Parameter pelatihan dikonfigurasi secara spesifik untuk mencerminkan karakteristik dataset, meliputi *network size*, jumlah kelas deteksi (dalam hal ini hanya satu kelas: "*Human*"), dan jumlah filter. Proses pelatihan dijalankan hingga mencapai 6.000 iterasi, berdasarkan hasil uji awal yang menunjukkan bahwa angka tersebut cukup untuk menghindari *overfitting* sekaligus mencapai kestabilan performa.

Selama pelatihan, performa model dievaluasi menggunakan parameter *mean Average Precision* (mAP), yang berfungsi sebagai indikator kualitas sistem deteksi objek. Metrik ini mencerminkan kemampuan model dalam mengenali dan mengklasifikasikan objek target secara akurat pada berbagai posisi dan skala dalam citra. Evaluasi dilakukan setiap 1.000 iterasi, dan hasilnya disimpan secara otomatis untuk dianalisis. Nilai mAP tertinggi yang dicapai merepresentasikan performa terbaik dan selanjutnya digunakan pada tahap pengujian untuk mendeteksi objek.

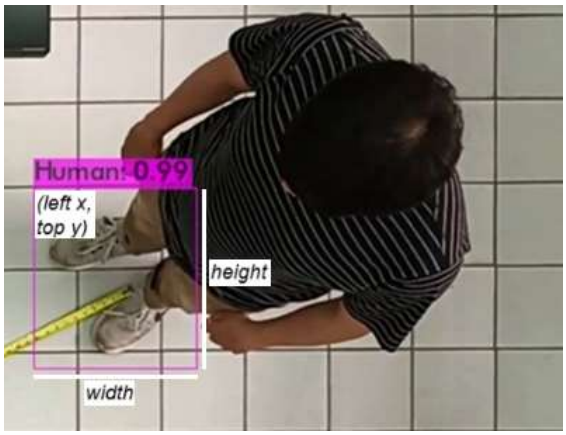
2.4. Deteksi Posisi Objek

Setelah proses pelatihan, model diterapkan untuk mendeteksi posisi objek pada citra uji. Citra uji tersebut diambil menggunakan kamera dengan lokasi berbeda dari dataset yang digunakan selama pelatihan. Deteksi objek dilakukan dengan mengidentifikasi *bounding box* kaki dan menghitung titik tengah antar kaki. Titik ini merepresentasikan posisi objek dan selanjutnya dikonversi dari koordinat pixel ke satuan centimeter berdasarkan rasio konversi yang diperoleh dari data kalibrasi.

Perhitungan posisi objek dilakukan dengan mengacu pada titik tengah citra sebagai referensi. Titik ini ditentukan dari ukuran resolusi gambar yang digunakan, yaitu 1280×720 pixel, sehingga koordinat referensinya berada pada (640, 360), seperti digambarkan pada Gambar 3. Setiap hasil deteksi dari model menghasilkan parameter *bounding box* berupa *left_x*, *top_y*, *width*, dan *height,* yang ditunjukkan pada Gambar 4. Parameter *left_x* dan *top_y* menyatakan koordinat pixel dari titik pojok kiri atas pada *bounding box*. Sedangkan *width* dan *height* merepresentasikan dimensi horizontal dan vertikal dari *bounding box* tersebut dalam satuan pixel. Titik tengah dari *bounding box* dihitung dengan persamaan (1) berikut:

$$\textbf{Koordinat titik tengah} = \left(x_{left} + \frac{width}{2}, y_{top} + \frac{height}{2}\right) \quad (1)$$

Gambar 3. Representasi titik tengah citra (640, 360) pada gambar beresolusi 1280×720 pixel sebagai acuan perhitungan posisi. Titik (0, 0) berada di kiri atas, sedangkan (1280, 720) di kanan bawah.



Gambar 4. *Struktur bounding box* hasil deteksi objek: *left_x* dan *top_y* menunjukkan koordinat kiri atas, *width* dan *height* menunjukkan dimensi horizontal dan vertikal kotak deteksi.

Selanjutnya, selisih antara titik tengah deteksi dan titik tengah citra dihitung dengan metode pengurangan absolut, sehingga hasil koordinat terpusat pada titik (0, 0) dengan menggunakan persamaan (2) dan (3) sebagai beriktu:

$$\boldsymbol{X(x) = x - 640} \quad (2)$$

$$\boldsymbol{Y(y) = y - 360} \quad (3)$$

Nilai (X, Y) ini menyatakan selisih posisi objek dari titik tengah gambar dalam satuan pixel. Untuk mengubahnya ke dalam satuan centimeter, dilakukan proses kalibrasi terhadap data citra yang diambil menggunakan kamera *smartphone* dengan posisi tetap. Perhitungan kalibrasi berdasarkan perbandingan antara ukuran nyata area citra dengan ukuran pixel. Hasil pengukuran menunjukkan bahwa area pengamatan memiliki dimensi fisik 395,0 cm (panjang) dan 221,8 cm (lebar), sedangkan dimensi citra adalah 1280 pixel (panjang) dan 720 pixel (lebar). Maka diperoleh rasio konversi sebagai berikut:

$$\text{Rasio konversi sb. X} = \frac{395{,}0 \text{ cm}}{1280 \text{ pixel}} \approx 0.309 \text{ cm/pixel}$$

$$\text{Rasio konversi sb. Y} = \frac{221{,}8 \text{ cm}}{720 \text{ pixel}} \approx 0.308 \text{ cm/pixel}$$

Setelah jarak dalam pixel dihitung, digunakan persamaan (4) untuk memperoleh jarak total dari titik tengah deteksi ke titik tengah citra:

$$\textbf{Jarak (px)} = \sqrt{(x_{px} - \mathbf{640})^2 + (y_{px} - \mathbf{360})^2} \quad (4)$$

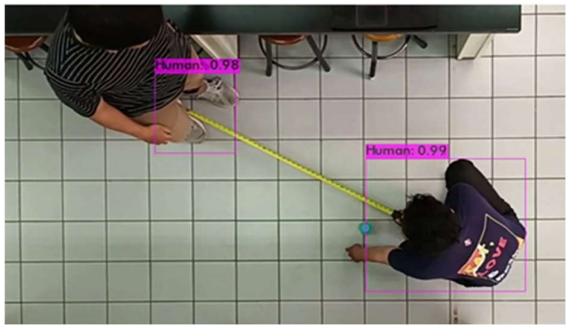
Jarak tersebut kemudian dikonversi ke centimeter menggunakan persamaan (5) berikut:

$$\textbf{Jarak (cm)} = \sqrt{(x_{cm} - \mathbf{197,5})^2 + (y_{cm} - \mathbf{110,9})^2} \quad (5)$$

2.5. Evaluasi Akurasi Deteksi

Evaluasi akurasi deteksi dilakukan dengan membandingkan jarak hasil deteksi terhadap jarak aktual sejauh 100 cm, seperti ditunjukkan pada Gambar 5. Perhitungan *error* dilakukan dengan menggunakan rumus:

$$\textbf{Error}(\%) = \left|\frac{Jarak_{deteksi} - Jarak_{aktual}}{Jarak_{aktual}}\right| \times \mathbf{100}\% \quad (6)$$



Gambar 5. Ilustrasi pengambilan data jarak aktual sejauh 100 cm sebagai acuan evaluasi akurasi sistem deteksi posisi.

Pengujian dilakukan terhadap sebelas citra uji dengan variasi posisi objek, yaitu di pojok kiri atas, kanan atas, kiri bawah, kanan bawah, dan tengah. Akurasi hasil deteksi kemudian dibandingkan dengan kebutuhan presisi antara 1–10 cm. Sistem dinilai layak digunakan jika *error* berada dalam rentang tersebut. Selain itu, hasil deteksi objek dibandingkan dengan hasil penelitian sebelumnya, baik dari sisi akurasi deteksi, maupun kemampuannya dalam deteksi terhadap gambar uji yang tidak dikenali sebelumnya.

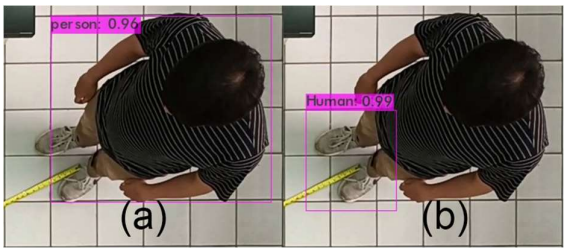
III. HASIL DAN PEMBAHASAN

Hasil proses pelatihan model menunjukkan peningkatan nilai *mean Average Precision* (mAP) seiring jumlah iterasi, sebagaimana ditunjukkan pada Tabel 1. Nilai mAP pada setiap 1000 iterasi pelatihan model.Tabel 1. Model mencapai mAP maksimum sebesar 100% setelah 4000 iterasi, dengan rata-rata mAP sebesar 99,19%. Ini menunjukkan bahwa model mampu mengenali objek manusia dengan sangat baik pada citra tampak atas.

Tabel 1. Nilai mAP pada setiap 1000 iterasi pelatihan model.

No	Jumlah Iterasi	mAP
1	1000	95,52%
2	2000	99,90%
3	3000	99,99%
4	4000	100,00%
5	5000	100,00%
6	6000	100,00%
Rata-rata		**99,19%**

Model yang telah dilatih kemudian digunakan untuk mendeteksi posisi objek pada citra uji. Hasil deteksi objek ditunjukkan pada Gambar 6, yang memperlihatkan perbandingan antara pendekatan studi sebelumnya dan metode pada penelitian ini. Deteksi sebelumnya menghasilkan *bounding box* dengan *confidence score* 0,96, sedangkan pendekatan baru menghasilkan *bounding box* pada kaki dengan *confidence score* lebih tinggi, yaitu 0,99, sehingga lebih sesuai untuk perhitungan posisi berdasarkan titik tengah antar kaki.



Gambar 6. Perbandingan hasil deteksi objek: (a) metode sebelumnya dengan satu *bounding box* umum, (b) metode anotasi penelitian ini dengan *bounding box* pada area kaki yang menghasilkan deteksi *confidence score* lebih tinggi.

Selanjutnya, perhitungan posisi objek dilakukan berdasarkan titik tengah dari *bounding box* yang dihasilkan oleh proses deteksi. Hasil analisis, sebagaimana ditunjukkan pada Tabel 2, menunjukkan bahwa rata-rata *error* posisi adalah sebesar 2,78 cm dengan standar deviasi sebesar 1,42 cm, yang setara dengan 2,78% dari jarak aktual sejauh 100 cm.

Tabel 2. Analisis error estimasi posisi hasil deteksi

No. Citra Uji	Hasil Estimasi Posisi		Error (cm)	Error (%)
	Jarak (px)	Jarak (cm)		
1	340,56	105,06	5,06	5,06
2	317,08	97,81	2,19	2,19
3	319,09	98,44	1,56	1,56
4	316,74	97,71	2,29	2,29
5	313,75	96,79	3,21	3,21
6	333,23	102,83	2,83	2,83
7	335,33	103,48	3,48	3,48
8	328,69	101,43	1,43	1,43
9	341,78	105,46	5,46	5,46
10	328,99	101,46	1,46	1,46
11	319,12	98,42	1,58	1,58
		Rata-rata Error		**2,78**
		Standar deviasi		**1,42**

Dibandingkan dengan hasil penelitian sebelumnya yang memiliki *error* sebesar 40,14% dan standar deviasi 4,58 cm [17], sistem ini menunjukkan peningkatan akurasi sebesar 93,07% serta penurunan standar deviasi sebesar 68,99%. Hal ini mengindikasikan peningkatan yang signifikan dalam akurasi dan konsistensi sistem deteksi posisi yang dikembangkan.

Peningkatan akurasi diperoleh karena posisi objek dihitung berdasarkan titik tengah antara kedua kaki. Pendekatan ini berbeda dari penelitian sebelumnya yang melakukan kalibrasi berdasarkan titik tengah dada. Titik dada dinilai kurang konsisten karena perbedaan tinggi badan antar individu dan memerlukan penyesuaian tambahan agar citra akurat.

Sebaliknya, titik tengah kaki dianggap lebih stabil, mudah dianotasi, dan dekat dengan titik tengah gambar. Meskipun kaki tidak selalu terlihat, posisinya tetap dapat diperkirakan dengan bantuan garis perspektif, seperti menarik garis dari ketiak menuju titik hilang citra.

IV. KESIMPULAN

Penelitian ini mengembangkan metode anotasi visual baru pada sistem deteksi posisi objek dalam ruangan menggunakan algoritma YOLOv4 yang diimplementasikan pada citra dari kamera *smartphone*. Model yang dilatih mencapai nilai mAP rata-rata sebesar 99,19%, dengan performa deteksi maksimal pada iterasi ke-4000 hingga 6000 sebesar 100%.

Perhitungan posisi objek dilakukan berdasarkan titik tengah antar kaki dari *bounding box* deteksi,

yang kemudian dikonversi ke satuan nyata menggunakan rasio kalibrasi. Hasil evaluasi menunjukkan rata-rata error sebesar 2,78 cm dengan standar deviasi 1,42 cm, atau setara dengan 2,78% dari jarak aktual.

Dibandingkan dengan studi sebelumnya, pendekatan ini menghasilkan peningkatan akurasi sebesar 93,07% dan penurunan deviasi sebesar 68,99%, menandakan sistem yang lebih presisi dan stabil. Sistem ini dapat digunakan untuk aplikasi pelacakan posisi di dalam ruangan sempit, khususnya dalam konteks pemantauan atau navigasi berbasis kamera tanpa sensor tambahan.

Untuk pengembangan penelitian selanjutnya, variasi dataset dan skenario pengujian dapat diperluas, mencakup beragam kondisi pencahayaan, latar belakang, serta sudut pandang kamera. Jika model akan digunakan pada berbagai jenis kamera, maka pembuatan dataset yang lebih besar menjadi penting guna meminimalkan risiko *overfitting*, terutama jika tetap menggunakan anotasi titik tengah di antara dua kaki.

UCAPAN TERIMA KASIH

Terima kasih kepada KBK Fisika Instrumentasi Program Studi Fisika FPMIPA UPI yang sudah memfasilitasi penelitian ini.

REFERENSI

[1] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019, doi: 10.1109/COMST.2019.2911558.

[2] C. J. N. Syazwani, N. H. A. Wahab, N. Sunar, S. H. S. Ariffin, K. Y. Wong, and Y. Aun, "Indoor Positioning System: A Review," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 477–490, Autumn 2022, doi: 10.14569/IJACSA.2022.0130659.

[3] J. Kunhoth, A. G. Karkar, S. Al-Maadeed, and A. Al-Ali, "Indoor positioning and wayfinding systems: a survey," *Human-centric Computing and Information Sciences 2020 10:1*, vol. 10, no. 1, pp. 1–41, May 2020, doi: 10.1186/S13673-020-00222-0.

[4] Y. Shi *et al.*, "Design of a Hybrid Indoor Location System Based on Multi-Sensor Fusion for Robot Navigation," *Sensors 2018, Vol. 18, Page 3581*, vol. 18, no. 10, p. 3581, Oct. 2018, doi: 10.3390/S18103581.

[5] J. Huang, S. Junginger, H. Liu, and K. Thurow, "Indoor Positioning Systems of Mobile Robots: A Review," *Robotics 2023, Vol. 12, Page 47*, vol. 12, no. 2, p. 47, Mar. 2023, doi: 10.3390/ROBOTICS12020047.

[6] J. B. Kim and H. S. Jun, "Vision-based location positioning using augmented reality for indoor navigation," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 954–962, 2008, doi: 10.1109/TCE.2008.4637573.

[7] B. C. Huang, J. Hsu, E. T. H. Chu, and H. M. Wu, "ARBIN: Augmented Reality Based Indoor Navigation System," *Sensors 2020, Vol. 20, Page 5890*, vol. 20, no. 20, p. 5890, Oct. 2020, doi: 10.3390/S20205890.

[8] F. Ahmed, M. Phillips, S. Phillips, and K. Y. Kim, "Comparative Study of Seamless Asset Location and Tracking Technologies," *Procedia Manuf*, vol. 51, pp. 1138–1145, Jan. 2020, doi: 10.1016/J.PROMFG.2020.10.160.

[9] S. J. Hayward, J. Earps, R. Sharpe, K. van Lopik, J. Tribe, and A. A. West, "A novel inertial positioning update method, using passive RFID tags, for indoor asset localisation," *CIRP J Manuf Sci Technol*, vol. 35, pp. 968–982, Nov. 2021, doi: 10.1016/J.CIRPJ.2021.10.006.

[10] A. Morar *et al.*, "A Comprehensive Survey of Indoor Localization Methods Based on Computer Vision," *Sensors 2020, Vol. 20, Page 2641*, vol. 20, no. 9, p. 2641, May 2020, doi: 10.3390/S20092641.

[11] K. Maulana Azhar, I. Santoso, D. Yosua, and A. A. Soetrisno, "IMPLEMENTASI DEEP LEARNING MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DAN ALGORITMA YOLO DALAM SISTEM PENDETEKSI UANG KERTAS RUPIAH BAGI PENYANDANG LOW VISION," *Transient: Jurnal Ilmiah Teknik Elektro*, vol. 10, no. 3, pp. 502–509, Sep. 2021, doi: 10.14710/TRANSIENT.V10I3.502-509.

[12] J. Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Cvpr*, vol. 2016-December, pp. 779–788, Dec. 2016, doi: 10.1109/CVPR.2016.91.

[13] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, Accessed: May 31, 2025. [Online]. Available: https://arxiv.org/pdf/2004.10934

[14] G. Kucukayan and H. Karacan, "YOLO-IHD: Improved Real-Time Human Detection System for Indoor Drones," *Sensors 2024, Vol. 24, Page 922*, vol. 24, no. 3, p. 922, Jan. 2024, doi: 10.3390/S24030922.

[15] J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," *J Phys Conf Ser*, vol. 1004, no. 1, p. 012029, Apr. 2018, doi: 10.1088/1742-6596/1004/1/012029.

[16] C. Liu, Y. Tao, J. Liang, K. Li, and Y. Chen, "Object detection based on YOLO network," *Proceedings of 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference, ITOEC 2018*, pp. 799–803, Dec. 2018, doi: 10.1109/ITOEC.2018.8740604.

[17] T. N. A. Kurniawan, "SISTEM DETEKSI POSISI OBJEK DALAM RUANGAN MENGGUNAKAN KAMERA SMARTPHONE DENGAN ALGORITMA COMPUTER VISION YOLO," Jan. 2024, Accessed: May 30, 2025. [Online]. Available: https://perpustakaan.upi.edu/nidn