



PENERAPAN DEEP LEARNING YOLO UNTUK DETEKSI MANUSIA

H Herdianto^{1*}, Darmeli Nasution², Hafni Hafni³, Syahrul Ramadhan⁴

^{1,2,3,4} Program Studi Sistem Komputer, Universitas Pembangunan Panca Budi, Medan

e-mail: *¹herdianto@dosen.pancabudi.ac.id, ²darmelinasution@dosen.pancabudi.ac.id,
³hafni@dosen.pancabudi.ac.id, ⁴syahrulramadhan5918@gmail.com

*Corresponding Author

Abstrak

Berdasarkan data yang diperoleh dari pihak kepolisian masalah pencurian sepeda motor masih cukup tinggi dan dapat saja terjadi di berbagai tempat seperti perparkiran, sekolah, rumah, kantor dan lain sebagainya dan pelakunya adalah laki – laki dengan bentuk dan tingginya juga berbeda-beda. Selain itu modusnya juga bervariasi dari mulai posisi manusia duduk, berdiri serta jongkok. Secara umum ada beberapa metode yang telah dikembangkan untuk deteksi manusia seperti Template Matching, Background Subtraction, Viola Jones (VJ), Histogram Oriented Gradient (HOG) dan Metode Deformable Part-based Model (DPM) tetapi masih terdapat kekurangan yaitu pada posisi manusia, waktu deteksi dan akurasi. Oleh karena itu tujuan dari penelitian ini adalah untuk mengetahui tingkat akurasi dari metode YOLO dalam mendeteksi objek manusia dari berbagai posisi. Adapun langkah – langkah yang digunakan untuk menyelesaikan penelitian ini yaitu studi literature, mengumpulkan data citra objek manusia, menentukan data latih dan testing, merancang program YOLO, melakukan pelatihan, testing. Hasil pengujian dari metode YOLO belum diketahui karena masih dalam proses pengujian.

Kata kunci— deteksi objek; deep learning; YOLO; CNN; convolusi.

Abstract

Based on data obtained from the police, the problem of motorbike theft is still quite high and can occur in various places such as parking lots, schools, homes, offices and so on and the perpetrators are men with different shapes and heights. Apart from that, the mode also varies from human positions sitting, standing and squatting. In general, there are several methods that have been developed for human detection such as Template Matching, Background Subtraction, Viola Jones (VJ), Histogram Oriented Gradient (HOG) and the Deformable Part-based Model (DPM) method but there are still shortcomings, namely in human position, time detection and accuracy. Therefore, the aim of this research is to determine the level of accuracy of the YOLO method in detecting human objects from various positions. The steps used to complete this research are literature study, collecting human object image data, determining training and testing data, designing the YOLO program, conducting training, testing. The test results of the YOLO method are not yet known because it is still in the testing process.

Keywords— object detection; distance measurement; estimation; monocamera.

I. PENDAHULUAN

Berdasarkan Undang-Undang No.2 Tahun 2002 Pasal 13 tentang Tugas Pokok Kepolisian, disebutkan bahwa pihak kepolisian telah melakukan tindakan pencegahan (preventive), pendekatan

(preemptive), dan penegakan hukum (repressive). Maka berdasarkan UU No. 2 di atas tindakan pencegahan terhadap terjadinya kasus pencurian bisa dilakukan oleh calon korban maupun upaya dari pihak kepolisian. Meskipun berbagai upaya telah dilakukan untuk mencegah terjadinya kasus pencurian namun tindak kejahatan ini



masih saja terjadi di kawasan tempat tinggal atau perumahan masih banyak terjadi sehingga mendapat perhatian khusus (Delia, 2009). Kejahatan yang umumnya terjadi di lingkungan tempat tinggal berbentuk pencurian rumah kosong, pencurian kendaraan bermotor, pencurian dengan penipuan, perampokan, serta pembobolan rumah (breaking & entering). Untuk melakukan monitoring ruangan membutuhkan bantuan beberapa alat pendukung. Adapun alat yang paling banyak digunakan adalah Closed Circuit Television (CCTV) dan IP Camera. Beberapa kekurangan dari kedua alat tersebut adalah kebutuhan instalasi terlebih dahulu; memerlukan kabel, konektor, adaptor, dan material bantu lainnya saat instalasi (penambahan biaya); lebih mudah dipengaruhi interferensi dan bisa disadap; rekaman gerakan umumnya mengalami perlambatan atau delay bergantung pada panjang kabel (CCTV) atau spesifikasi kamera pada IP Camera. Hal tersebut menjadi hambatan serius apabila pengguna membutuhkan solusi monitoring dengan biaya lebih terjangkau, praktis dan instalasi mudah, serta membutuhkan sedikit daya. Dalam (Supiyandi et al., 2023), hingga saat ini, telah diciptakan beberapa perangkat monitoring ruangan atau keadaan sekitar dengan memanfaatkan perangkat IoT dan microcontroller. Perangkat tersebut antara lain ATmega328, Arduino 101, dan Raspberry Pi. Metode pendukung yang digunakan untuk melakukan monitoring juga bermacam-macam. Metode yang digunakan adalah dengan menggunakan sensor Passive Infra Red (PIR), lalu ada juga dengan cara mendeteksi panas yang dikeluarkan dari tubuh objek (heat sensor), ataupun mendeteksi seseorang dengan menggunakan Red Laser Pointer dan Light Dependent Resistor.

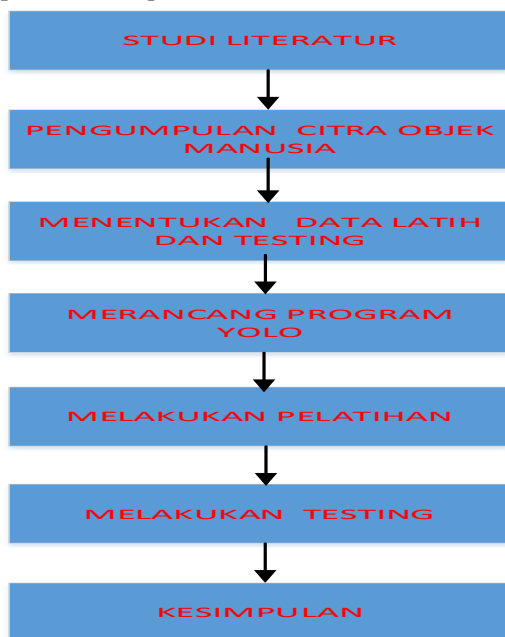
Salah satu penelitian monitoring ruangan adalah menerapkan pengolahan

citra dengan metode adaptive motion detection pada sistem kamera keamanan yang telah disertai push notification ke smartphone Android (Sikumbang & Suryadi, 2015),(Navneet & Triggs, 2005). Hasil penelitian menunjukkan bahwa angka false positive deteksi gerakan di dalam ruangan tergolong tinggi yakni 58,75%. Selain itu, tema object detection juga telah diteliti pada (Sikumbang & Suryadi, 2015),(Navneet & Triggs, 2005), (Herdianto, 2019),(Herdianto & Mursyidah, 2022),(Herdianto & Nasution, 2022), (Nababan et al., 2017), (Hariyanto et al., 2019). Pada (Sikumbang & Suryadi, 2015) dan (Navneet & Triggs, 2005) deteksi manusia dilakukan dengan menggunakan metode Histogram of Oriented Gradients berbasis OpenCV. Dalam penelitian tersebut, digunakan metode HOG sebagai metode deteksi objek berupa manusia dengan mencacah frame pada video menjadi frame gambar satu-persatu. Lalu pada penelitian (Herdianto, 2019) digunakan metode template matching dan background subtraction untuk mendeteksi manusia. Dari hasil penelitian ini dinyatakan bahwasannya metode template matching memiliki tingkat akurasi hingga 80% dalam mendeteksi objek manusia sedangkan background subtraction lebih cocok bila warna background citra berbanding terbalik dengan foreground. Selanjutnya ada juga penelitian lainnya yang dilakukan oleh (Herdianto & Mursyidah, 2022) tahun 2021. Untuk mendeteksi manusia khususnya bagian wajah digunakan metode viola jones dimana hasil metode ini dalam mendeteksi wajah manusia hingga 95%. Lalu ada juga penelitian tahun 2022 yang dilakukan oleh (Herdianto & Nasution, 2022) dengan menggunakan metode Convolutional Neural Network (CNN). Hasil penelitian ini menyatakan bahwasannya metode CNN dapat mengenali objek manusia dalam

berbagai bentuk dan posisi. Tetapi hasil penelitian ini masih memiliki kekurangan dalam hal waktu dalam deteksi objek yaitu 10 detik. Dari permasalahan di atas maka akan dilakukan sebuah penelitian untuk merancang sebuah sistem yang dapat digunakan untuk melakukan deteksi pergerakan objek berupa manusia dengan metode YOLO.

II. METODOLOGI PENELITIAN

Untuk menyelesaikan penelitian ini maka peneliti menyusun kerangka penelitian seperti Gambar 1.



Gambar 1. Kerangka Penelitian

Keterangan :

- Studi Literatur

Pada tahap ini peneliti mengumpulkan data terkait perkembangan hasil penelitian tentang deteksi objek manusia yang telah dilakukan peneliti sebelumnya. Data yang dikumpulkan bersumber dari prosiding, jurnal nasional dan internasional serta buku.

- Pengumpulan Citra Objek Manusia

Tahap ini peneliti mengambil sampel data citra manusia dengan berbagai bentuk, ukuran, posisi, warna kulit yang diperoleh

dari berbagai institusi seperti kaggle, KITTI dan lain sebagainya.

- Menentukan Data Latih dan Testing

Dari sekian banyak data sampel objek manusia yang diperoleh maka dilakukan pengelompokan menjadi 2 kategori yaitu data latih dan testing. Untuk data latih jumlahnya 70% dari data keseluruhan dan sisanya 30% untuk data testing

- Merancang Program YOLO

Untuk mengetahui apakah jaringan / arsitektur YOLO dapat mendeteksi manusia dengan benar maka ditanamkan program YOLO pada komputer.

- Melakukan Pelatihan

Agar computer dapat mendeteksi keberadaan manusia pada suatu citra dengan benar maka terlebih dahulu arsitektur YOLO yang telah dirancang dilakukan proses pelatihan hingga nilai error yang diinginkan telah tercapai atau waktu pelatihan terpenuhi.

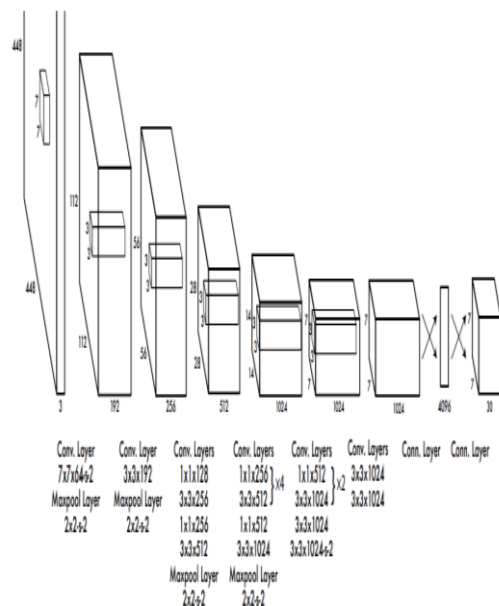
- Melakukan Testing

Selanjutnya dilakukan pengujian terhadap arsitektur YOLO yang telah selesai dilatih.

- Kesimpulan

Menarik kesimpulan dari hasil pengujian yang telah dilakukan untuk mengetahui ketercapaian tujuan penelitian.

YOLO merupakan salah satu metode deep learning yang dapat dipakai untuk mendeteksi berbagai objek salah satunya manusia. Pada YOLO berisi banyak layer dan merupakan proses convolusi. Antara layer awal dengan layer berikutnya saling berhubungan dan membentuk sebuah jaringan atau dikenal dengan istilah arsitektur. Ada beberapa arsitektur jaringan yang dipakai pada YOLO antara lain :CNN, darknet-19, darknet-53, Cross Stage Partial Network (CSP-Net), feature pyramid networks (FPN), EfficientNet, EfficientNet – L2.



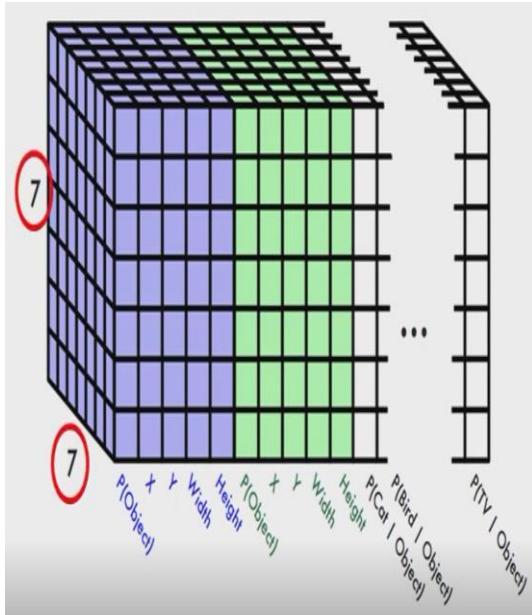
Gambar 2. Arsitektur Yolo (Redmon et al., 2016), (Redmon & Farhadi, 2017)

Gambar 2 merupakan salah satu arsitektur YOLO yang menggunakan CNN. Adapun cara kerja dari bagian - bagian arsitektur YOLO tersebut adalah sebagai berikut :

- layer 1 yang menjadi masukan citra pada YOLO dengan berbagai ukuran diubah ukurannya menjadi $448 * 448$. Lalu pada layer ini dilakukan proses konvolusi dengan ukuran windows $7*7$ dengan maxpooling layer 1 berukuran $2*2$. Proses ini dilakukan sebanyak 3 kali karena pada layer ini merupakan citra Red Green Blue (RGB).
- layer 2 juga masih merupakan proses konvolusi dimana masukan pada layer ini berukuran $112*112$ yang merupakan keluaran layer 1. Masukan pada layer ini dikonvolusi dengan menggunakan windows berukuran $3*3$ dan ada 192 kali proses konvolusi. Untuk menentukan keluaran layer ini digunakan maxpooling berukuran $2*2$.
- layer 3 juga merupakan proses konvolusi dimana masukan pada layer ini berukuran $56*56$ yang merupakan keluaran layer 2. Masukan pada layer ini dikonvolusi

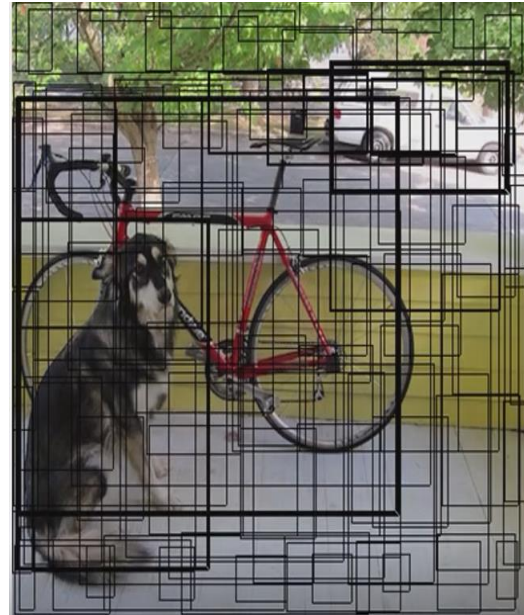
masih menggunakan windows yang berukuran 3*3 dan pada layer ini ada 256 kali proses konvolusi. Untuk menentukan keluaran layer ini digunakan maxpooling berukuran 2*2.

- layer 4 juga proses konvolusi dimana masukan pada layer ini lebih kecil dari masukan layer sebelumnya berukuran 28×28 yang merupakan keluaran layer 3. Masukan pada layer ini dikonvolusi dengan menggunakan windows berukuran 3×3 dan ada 512 kali proses konvolusi. Untuk menentukan keluaran layer ini digunakan maxpooling berukuran 2×2 .
- pada layer 5 masukannya menjadi lebih kecil yaitu berukuran 14×14 lalu dilakukan proses konvolusi sebanyak 1024 kali sehingga luaran dari layer ini menjadi lebih kecil dan berukuran 7×7 .
- layer 6 merupakan proses flatten dimana flatten adalah proses merubah dari bentuk keluaran citra matriks dua dimensi pada pooling layer menjadi satu kolom (sebuah vektor satu dimensi). Disini citra matriks yang berukuran $7 \times 7 \times 1024$ diubah menjadi 1 dimensi dengan ukuran 4096.
- layer 7 merupakan fully connected atau bentuk umum dari jaringan syaraf tiruan. Pada layer ini ukuran citra sudah menjadi $7 \times 7 \times 30$ seperti terlihat pada Gambar 3.

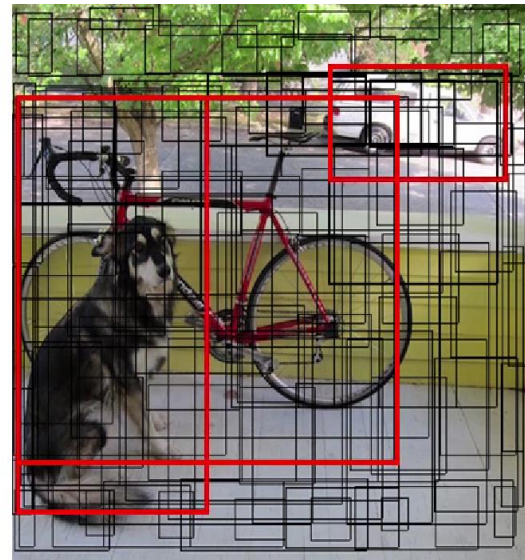


Gambar 3. keluaran YOLO (Lim et al., 2021)

Gambar 3 merupakan bentuk masukan dan keluaran YOLO. Dimana 7×7 merupakan lebar dan tinggi citra yang dapat diibaratkan memiliki 7 grid sedangkan 30 untuk jumlah kategori 20. Masing sel akan memprediksi 2 bounding box beserta confidencenya. Untuk confidence 1 ada di saluran 1, bounding box 1 ada di saluran 2 hingga 5, confidence ke 2 ada di saluran 6 untuk bounding boxnya ada di saluran 7 hingga 10. Saluran 11 hingga 30 merupakan lokasi kategori bila berjumlah 20. Bila setiap sel memiliki 2 prediksi bounding box maka kemungkinan hasil prediksinya bounding box berjumlah $7 \times 7 \times 2 = 98$. Bila digambarkan akan terlihat seperti Gambar 4.

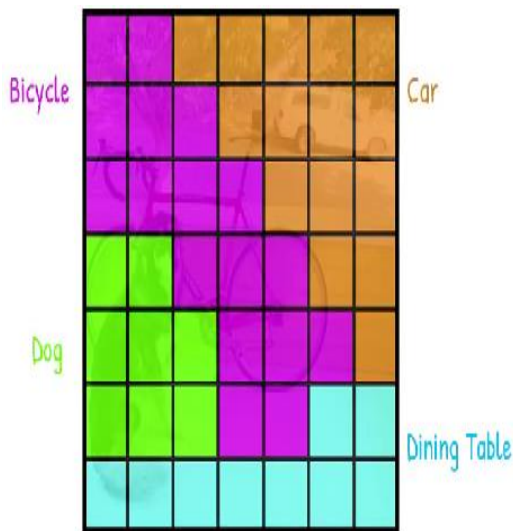


Gambar 4. total bounding box hasil prediksi



Gambar 5. bounding box dengan confidence terbesar

Dari 98 bounding box yang terbentuk maka untuk menentukan mana bounding box yang memiliki nilai confidence yang tertinggi dilihat dari ketebalan garis bounding box yang terbentuk pada Gambar 5 berwarna merah. Lalu nilai x , y , width dan height pada Gambar 3 dinormalisasi ke 0 hingga 1. Selanjutnya selain bounding box YOLO juga memprediksi kategori objek disetiap selnya seperti Gambar 6.



Gambar 6. hasil prediksi

Di bagian keluaran seperti Gambar 3. bila kategorinya ada 20 maka 20 saluran terakhir merupakan kategori objek artinya masing – masing box ada kategorinya. Lalu untuk menentukan bounding box yang ada objeknya digunakan nilai threshold dimana bila bounding box yang memiliki nilai confidence dan Non Maximum Suppresion (NMS) lebih tinggi dari nilai threshold yang ditetapkan maka bounding box tersebut yang dipilih sedangkan yang memiliki nilai confidence yang rendah diabaikan.

III. HASIL DAN PEMBAHASAN

Spesifikasi Perangkat Keras dan Lunak

Adapun spesifikasi perangkat keras dan lunak yang digunakan pada penelitian seperti terlihat pada Tabel 1 dan 2.

Tabel 1. perangkat keras yang digunakan

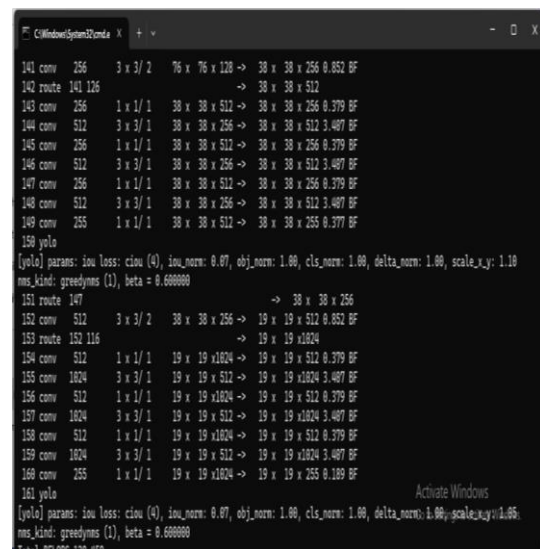
No	Nama perangkat
1	Prosesor core i5
2	RAM 16 giga byte
3	VGA GPU RTX 3060

Tabel 2. perangkat lunak yang digunakan

No	Nama perangkat
1	Visual studio code
2	Pytorch
3	Python
4	Cmake
5	Darknet

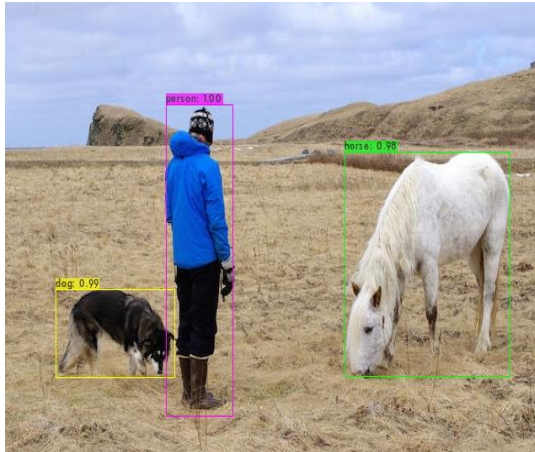
Pengujian

Pada penelitian ini untuk pengujian objek manusia belum dilakukan secara keseluruhan.



Gambar 7. Proses pelatihan arsitektur YOLO

Gambar 7. menunjukkan proses pelatihan arsitektur YOLO yang berisi layer – layer konvolusi. Dimana tujuan dari pelatihan ini adalah mengupdate nilai bobot pada setiap jaringan layer.



Gambar 8. hasil prediksi YOLO

Dari gambar 8. terlihat YOLO dapat mendeteksi setiap objek pada citra dengan memberi bounding box dan nama kategorinya dengan benar.

IV. KESIMPULAN

Hasil pengujian pada penelitian belum bisa disimpulkan karena masih dalam proses pengujian lebih lanjut.

V. SARAN

Untuk kesempurnaan penelitian ini maka peneliti mengusulkan agar dapat menggunakan arsitektur YOLO yang lain.

VI. DAFTAR PUSTAKA

- Delia, R. . (2009). Analisis Determinan Penyebab Timbulnya Fear of Crime Pada Kasus Pencurian Di Kalangan Ibu Rumah Tangga. *Indonesian Journal of Criminology*, 5(1), 1–5.
- Hariyanto, E., Iqbal, M., Siahaan, A. P. U., Saragih, K. S., & Batubara, S. (2019). Comparative study of tiger identification using template matching approach based on edge patterns. *Journal of Physics: Conference Series*, 1196(1). <https://doi.org/10.1088/1742-6596/1196/1/012025>
- Herdianto. (2019). Perbandingan Metode Template Matching dengan Background Substraction untuk Mendeteksi Objek Manusia. *Core IT*, 7(2), 28–33.
- Herdianto, H., & Mursyidah, M. (2022).

Deteksi Wajah Manusia Pada Image Sequence Menggunakan Background Subtraction Dan Haar Cascade Classifier. 7(1).

- Herdianto, H., & Nasution, D. (2022). Klasifikasi Objek Menggunakan Metode Convolutional Neural Network (CNN). *SNASTIKOM*, 1–8.
- Lim, J. S., Astrid, M., Yoon, H. J., & Lee, S. I. (2021). Small Object Detection using Context and Attention. *3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2021*, 181–186. <https://doi.org/10.1109/ICAIIIC51459.2021.9415217>
- Nababan, E. B., Iqbal, M., & Rahmat, R. F. (2017). Breast cancer identification on digital mammogram using Evolving Connectionist Systems. *2016 International Conference on Informatics and Computing, ICIC 2016*, 132–136. <https://doi.org/10.1109/IAC.2016.7905703>
- Navneet, D., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1–8. https://doi.org/10.1007/978-3-642-33530-3_8
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua, 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>
- Sikumbang, S., & Suryadi, K. (2015). Human Detection Menggunakan



Metode Histogram Of Oriented Gradients (Hog) Berbasis Open CV. *Jurnal Pendidikan Teknik Elektro*, 4(2), 1–6.

Supiyandi, S., Rizal, C., Iqbal, M., Siregar, M. N. H., & Eka, M. (2023). Smart Home Berbasis Internet of Things (IoT) Dalam Mengendalikan dan Monitoring Keamanan Rumah. *JOURNAL OF INFORMATION SYSTEM RESEARCH (JOSH)*, 99(99), 1–7.
<https://doi.org/10.30865/jurikom.v9i1.9999>