

## IMPLEMENTASI METODE YOLO PADA DETEKSI OBJEK MANUSIA

**Herdianto<sup>✉</sup>, Hafni, Darmeli Nasution, Syahrul Ramadhan**

Program Studi Sistem Komputer, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: [herdianto@dosen.pancabudi.ac.id](mailto:herdianto@dosen.pancabudi.ac.id)

DOI: <https://doi.org/10.46880/jmika.Vol8No2.pp234-240>

### ABSTRACT

*Until now, the problem of theft of motorbikes and livestock in North Sumatra is still quite high. Motorbike theft can occur in many locations, including housing, schools, parking lots, offices and many other places, while for livestock it can occur on pastures and in pens during the day or night with the perpetrators being men. To make this theft a success, various modes are used in varying human positions, from sitting, squatting to standing. To help overcome this, several object detection methods have been developed such as Background Subtraction, Template Matching, Histogram Oriented Gradient and Viola Jones (VJ). Of the many methods that have been used, there are still shortcomings, namely in time, accuracy and various human positions. For this reason, research was carried out with the aim of improving the time and level of accuracy in detecting human objects using the YOLO method. From the trials carried out, it is known that YOLO can detect humans in various positions with a mAP value of 0.99 and an average detection time of 810.01 ms.*

**Keyword: Humans, Convolution, Object Detection, Yolo, Deep Learning.**

### ABSTRAK

*Sampai saat ini masalah pencurian seperti sepeda motor, hewan ternak di Sumatera Utara masih cukup tinggi. Untuk lokasi pencurian sepeda motor dapat terjadi di banyak lokasi antara lain perumahan, sekolah, perparkiran, kantor dan masih banyak tempat lainnya sedangkan untuk hewan ternak dapat terjadi lahan padang rumput dan di kandang pada siang maupun malam hari dengan pelakunya adalah laki-laki. Untuk mensukseskan aksi pencurian ini berbagai modus diterapkan dengan bervariasi posisi manusia dari mulai duduk, jongkok hingga berdiri. Untuk menyelesaikan persoalan tersebut di atas banyak metode deteksi objek telah dikembangkan seperti Background Subtraction, Template Matching, Histogram Oriented Gradient dan Viola Jones (VJ). Dari banyak metode yang telah dikerjakan masih ada kekurangan yaitu pada waktu, akurasi dan berbagai posisi manusia. Untuk itu dilakukan penelitian yang tujuan memperbaiki waktu dan tingkat akurasi dalam mendeteksi objek manusia menggunakan metode YOLO. Dari uji coba yang dilakukan diketahui YOLO dapat mendeteksi manusia pada berbagai posisi dengan nilai mAP sebesar 0,99 dan rata-rata waktu deteksi 810,01 ms.*

**Kata Kunci: Manusia, Convulasi, Deteksi Objek, Yolo, Deep Learning.**

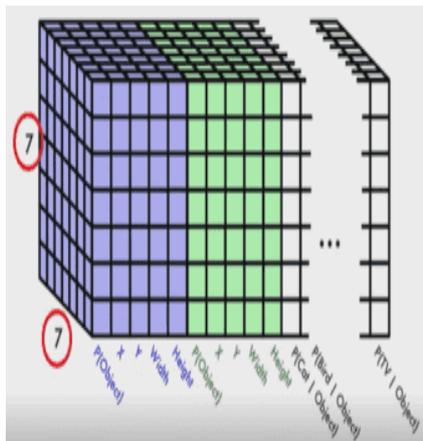
### PENDAHULUAN

Polisi merupakan aparat negara yang mempunyai tugas melakukan penindakan hukum, pencegahan dan pendekatan terkait tindak kejahatan yang telah dan akan dilakukan seseorang terhadap orang lain. Hal ini tertuang pada Undang-Undang No.2 pasal 13 Tahun 2002 tentang Tugas Pokok Kepolisian. Dari uraian singkat UU No. 2 tersebut maka tindakan ataupun upaya pencegahan terhadap tindak kejahatan seperti (pencurian) yang sifatnya berulang kembali dapat dilakukan secara swadaya artinya dilakukan oleh individu-individu sendiri ataupun berkolaborasi dengan pihak berwajib yaitu kepolisian. Telah banyak tindakan pencegahan yang telah diterapkan dengan tujuan mencegah tindakan pencurian tetapi ironisnya aktivitas kejahatan ini terus ada dan pada umumnya terjadi di kawasan perumahan atau tempat tinggal

lokal (Delia, 2009). Berbagai upaya swadaya masyarakat telah dilakukan untuk mencegah aksi pencurian antara lain melakukan monitoring daerah sekitar dengan memasang beberapa Closed Circuit Television (CCTV) di titik yang rawan. Akan tetapi implementasi dari upaya tersebut memiliki beberapa kekurangan yaitu banyaknya komponen instalasi yang dipakai seperti adaptor, konektor, memerlukan kabel yang cukup panjang sehingga menambah biaya pemasangan. Selain itu cara seperti ini mudah terjadi penyadapan, terjadi perlambatan pengiriman gambar karena pengaruh panjangnya kabel (CCTV) yang digunakan dan rentan gangguan interferensi sehingga berdampak pada hasil rekaman yang kurang baik. Ini menjadi tantangan tersendiri yang memerlukan penanganan serius jika pengguna bermaksud menggunakan monitoring dengan biaya ekonomis,



- lapisan empat masih terjadi konvolusi dengan masukan lapisan ini menjadi lebih sedikit dari masukan pada lapisan sebelumnya dan memiliki ukuran 28\*28 yang merupakan luaran lapisan tiga. Masukan lapisan ini di konvolusi memakai windows ukuran 3\*3 lalu dilakukan 512 konvolusi. Untuk menetapkan luaran pada lapisan ini dipakai max pooling ukuran 2\*2.
- pada lapisan lima masukan lebih kecil dari lapisan empat yang memiliki ukuran 14\*14 selanjutnya proses konvolusi dijalankan hingga 1024 kali akibatnya luaran lapisan lima ini menjadi semakin kecil dengan ukuran 7 \* 7.
- lapisan enam adalah proses flatten yang berfungsi merubah bentuk citra matriks dua dimensi di lapisan pooling menjadi sebuah vektor satu dimensi. Pada lapisan ini citra matriks dengan ukuran 7 \* 7 \* 1024 telah diubah menjadi satu dimensi berukuran 4096.
- lapisan tujuh adalah bentuk dari jaringan syaraf tiruan pada umumnya atau fully connected. Pada lapisan ini citra telah berubah ukuran menjadi 7 \* 7 \*30 seperti Gambar 2.

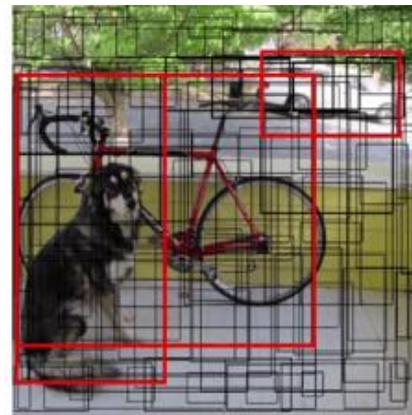


Gambar 2. YOLO Susunan Luaran (Lim, 2021)

Gambar 2 adalah masukan dan luaran YOLO. Untuk 7 \* 7 menyatakan tinggi dan lebar masukan citra diumpamakan mempunyai 7 grid dan 30 adalah kategorinya ada 20. Setiap sel dapat memprediksi sebanyak dua bounding box berikut confidencenya. Pada confidence satu terdapat di channel satu, bounding box satu pada channel dua sampai lima, sedangkan confidence ke dua terdapat di channel enam dengan bounding boxnya di channel tujuh sampai sepuluh. Channel sebelas hingga tiga puluh adalah alamat kategori jika ada dua puluh. Bila masing-masing sel mempunyai dua kemungkinan bounding box akibatnya jumlah bounding box yang dapat terbentuk sebanyak 98 diperoleh dari 7\*7\*2 diilustrasikan pada Gambar 3.

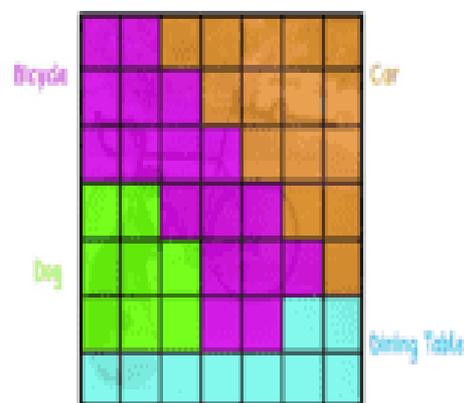


Gambar 3. Banyaknya Bounding Box Terbentuk



Gambar 4. Confidence Terbesar pada Bounding Box

Diketahui jumlah bounding box terbentuk ada 98 maka cara yang digunakan dalam menetapkan bounding box dengan nilai confidence terbesar ditentukan dari tebalnya garis yang membentuk bounding box seperti Gambar 5 diberi warna merah. Selanjutnya nilai y, x dan height serta width dari Gambar 2 dilakukan proses normalisasi ke nilai 0 sampai 1. Selain itu bounding box yang terbentuk pada YOLO juga dapat memprediksi banyaknya kategori dimasing-masing selnya diilustrasikan Gambar 5.



Gambar 5. Hasil Prediksi

Di bagian luaran terlihat pada Gambar 2. jika kategorinya berjumlah 20 artinya 20 channel terakhir adalah kategori objek yang artinya pada setiap box ada memiliki kategori. Selanjutnya untuk menetapkan bounding box yang terdapat objek dipakai sebuah nilai *threshold* dengan ketentuan jika bounding box dengan nilai Non Maximum Suppression dan confidence lebih besar dibandingkan nilai *threshold* maka diputuskan bounding box ini yang diambil dan bounding box dengan nilai confidence kecil tidak dianggap.

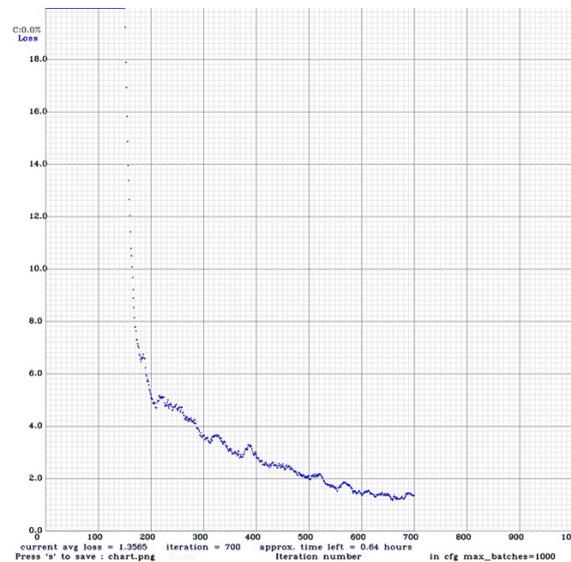
## METODE PENELITIAN

Adapun kerangka penelitian yang disusun pada penelitian ini sesuai tahapan berikut:

- Studi Pustaka  
Tahap ini dikumpulkan data dari berbagai sumber seperti jurnal nasional, prosiding nasional dan internasional serta buku terkait perkembangan penelitian deteksi objek manusia
- Pengumpulan Citra Manusia  
Peneliti mengambil dan mengumpulkan sampel citra manusia dengan bermacam-macam ukuran, warna kulit, bentuk, posisi yang didapat dari bermacam-macam website yaitu KITTI, kaggle dan lain-lain.
- Menentukan Data Untuk Latih serta Testing  
Sampel data manusia yang telah didapat selanjutnya dikelompokkan menjadi dua bagian yaitu data testing dan latih. Pada data latih memiliki persentase hingga 70% data keseluruhan sedangkan testing 30%.
- Merancang Program YOLO  
Agar jaringan/arsitektur YOLO menjadi pintar dapat mengenali objek manusia dengan baik maka pada komputer diberi kecerdasan buatan berupa program YOLO.
- Melakukan Pelatihan  
Sebelum arsitektur YOLO diterapkan maka terlebih dahulu dilakukan pelatihan untuk merubah dari pada bobot jaringan YOLO. Pelatihan ini akan berhenti bila waktu pelatihan terpenuhi atau nilai error yang diharapkan telah tercapai.
- Melakukan Testing  
Setelah pelatihan selesai dilanjutkan dengan tahap testing terhadap arsitektur YOLO apakah benar-benar arsitektur YOLO dapat mengenali objek sesuai data latih.
- Kesimpulan  
Dari hasil pengujian akan diketahui ketercapaian tujuan penelitian.

## HASIL DAN PEMBAHASAN

Adapun perbedaan penelitian ini dengan penelitian lainnya dalam mendeteksi objek manusia yaitu bila penelitian sebelumnya hanya wajah dan hanya memberikan label objek manusia. Pada YOLO dapat memberikan bounding box pada setiap objek manusia pada berbagai posisi. Agar jaringan YOLO dapat menjadi pintar dalam mengenali objek maka langkah awal yang dilakukan adalah melakukan pelatihan terhadap citra objek yang telah dikumpulkan. Dan tujuan dari pelatihan ini adalah untuk memperoleh bobot jaringan yang ideal yang nantinya akan dipakai pada proses testing. Jaringan YOLO yang digunakan dalam proses pelatihan dimana terlihat jaringan ini terdiri dari beberapa layer yang berisi konvolusi dengan berbagai filter.



Gambar 6. Proses pelatihan YOLO

Proses pelatihan pada gambar 6 menunjukkan perubahan nilai average loss (AVG) loss, iterasi, estimasi waktu pelatihan dan maksimum batches pelatihan. Saat pertama kali dilakukan pelatihan nilai AVG di atas 18, nilai ini masih tinggi jika untuk diterapkan. Selanjutnya nilai AVG ini terus diperbaiki hingga 0,98 dan nilai dianggap sudah ideal artinya jaringan sudah baik.

Untuk membuktikan bahwasannya jaringan hasil pelatihan sudah baik maka dilakukan proses pengujian dengan data citra yang pernah dilatih dan belum dengan persentase 70% data latih dan 30% belum pernah dilatihkan.



**Gambar 7.** YOLO Mendeteksi Objek

Pada gambar 7 memperlihatkan YOLO dapat mendeteksi objek manusia pada duduk ditandai dengan berhasilnya YOLO membentuk bounding box warna merah pada objek manusia tersebut.



**Gambar 9.** YOLO Mendeteksi Tiga Objek Manusia

Gambar 9 merupakan pengujian YOLO yang dapat mendeteksi tiga objek manusia dengan nilai mAP masing-masing 0,98 dan ditandai dengan berhasilnya YOLO membentuk bounding box warna merah pada tiga objek manusia.



**Gambar 8.** YOLO Mendeteksi Dua Objek Manusia

Pada gambar 8 YOLO dapat mendeteksi dua objek manusia pada posisi berdiri dengan nilai mean average precision (mAP) masing-masing 0,94 dan 0,99 dan ditandai dengan berhasilnya YOLO membentuk bounding box warna merah pada dua objek manusia tersebut.



**Gambar 10.** YOLO Mendeteksi Empat Objek Manusia

Gambar di atas merupakan pengujian metode YOLO dapat mendeteksi empat objek manusia pada posisi berdiri dan duduk dengan nilai mean average precision (mAP) masing-masing 0,85, 0,92, 0,99 dan 1,00 serta ditandai dengan berhasilnya YOLO membentuk bounding box warna merah pada setiap objek manusia tersebut. Waktu yang dibutuhkan YOLO dapat mendeteksi empat objek manusia sebesar 795,779 ms.



**Gambar 11.** YOLO Mendeteksi Lima Objek Manusia

Pada gambar di atas YOLO dapat mendeteksi lima objek manusia pada posisi berdiri, duduk dan berbagai tinggi dengan nilai mean average precision (mAP) masing- masing 0,81 dan lainnya 0,99 yang ditandai dengan berhasilnya YOLO membentuk bounding box warna merah pada lima objek manusia tersebut.

**Tabel 1.** Waktu Deteksi Objek Manusia

No	Objek Deteksi	Jumlah Objek Manusia	Waktu (ms)
1	Manusia	1	788,10
2	Manusia	2	793,55
3	Manusia	3	794,13
4	Manusia	4	795,77
5	Manusia	5	878,50

Tabel di atas menjelaskan terkait waktu yang dibutuhkan YOLO dalam mendeteksi objek manusia ketika berjumlah 1 dibutuhkan waktu selama 788,10 ms untuk jumlah 2 dibutuhkan waktu selama 793,55 ms dan seterusnya.

## KESIMPULAN

Selama penelitian dilakukan beberapa pengujian dari pengujian tersebut diketahui deep learning YOLO dapat mendeteksi objek manusia dan membentuk bounding box dengan nilai mAP 0,99 dan waktu deteksi rata-rata 810,01 ms.

## DAFTAR PUSTAKA

Anin, Alex, K., Ilya, S., & E Geoffrey, H. (2012). Imagenet classification with deep convolutional neural networks. *NIPS Conference*, 1097–1105.  
 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.  
 Felzenszwalb, P., B. Girshick, R., McAllester, D., & Ramanan, D. (2009). Object Detection with Discriminatively Trained Part Based Models. *Computer*, 47(2), 1–19.  
 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 Subtraction 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.  
 Herdianto, H., & Nasution, D. (2023). Implementasi Metode Cnn Untuk Klasifikasi Objek. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 7(1), 54–60.  
<https://doi.org/10.46880/jmika.vol7no1.pp54-60>  
 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/ICAIC51459.2021.9415217>  
 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, Iaic*, 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](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.
- Viola, P., & Jones, M. (2001). Rapid Object Detection  
using a Boosted Cascade of Simple Features.  
*Conference On Computer Vision And Pattern  
Recognition*, 1–9.
- Viola, P., & Jones, M. (2004). Robust Real-Time Face  
Detection Intro to Face Detection. *International  
Journal of Computer Vision*, 57(2), 137–154.