

# HUMAN ACTIVITY RECOGNATION DARI REKAMAN VIDEO PENGAWAS DENGAN METODE YOLO

Harry Ronaldo Yudistira<sup>1</sup>, Lina<sup>2</sup>

<sup>1,2)</sup> Teknik Informatika, FTI, Universitas Tarumanagara  
Jl. Letjen S Parman no 1, Jakarta 11440 Indonesia  
email : <sup>1</sup>[harryronaldo93@gmail.com](mailto:harryronaldo93@gmail.com) , <sup>2</sup>[lina@fti.untar.ac.id](mailto:lina@fti.untar.ac.id)

## ABSTRACT

*Advances in technology today, so that human work can be helped easily by using machines. One of the jobs that can be time-consuming is supervising. The supervision carried out requires human labor to oversee the events and activities recorded on CCTV cameras. With the development of technology, conducting surveillance is now easier. By utilizing video recordings containing activities, the design made can produce a program that can detect Sitting, Standing, Studying, Raising Hands, and Clapping activities. These detections can then be summarized into an activity logbook along with the time the activity occurred. The system is made using Python and the You Only Look Once (YOLO) method. The program is expected to be able to accurately detect activities with the specified classes. The results show that the YOLO method can find objects and their activities using the internet dataset and IP Camera dataset which produces the highest mean Average Precision (mAP) of 86.85% for the internet dataset and 99.96% for the IP Camera dataset. And the test results on the best model show an accuracy rate of 80.6% for the internet dataset and 98% for the IP Camera dataset.*

## Key words

*Activity Logbook, Activity Recording Video, Human Activity Recognition, Supervision, You Only Look Once*

## 1. Pendahuluan

Pada jaman sekarang teknologi semakin maju dan berkembang sehingga kemajuan teknologi dapat menggantikan aktivitas dan pekerjaan manusia dalam berbagai bidang contohnya seperti monitoring. Monitoring adalah aktivitas yang bertujuan untuk mengawasi perkembangan yang terjadi ataupun

mengamati berbagai suatu aktivitas yang terjadi disuatu tempat. Monitoring biasanya dilakukan disuatu tempat tertentu untuk mendokumentasikan hasil kejadian yang terjadi di tempat tersebut. Aktivitas ini memiliki masalah dan keterbatasan karena monitoring harus dilakukan secara manual ataupun secara langsung oleh tenaga kerja manusia.

Closed Circuit Television (CCTV) adalah kamera video yang berfungsi untuk mentransmisikan video ke tempat yang spesifik [1]. CCTV memiliki tujuan untuk memonitor situasi dan kondisi pada tempat tertentu sebagai bukti suatu kejadian yang telah terjadi pada waktu tertentu. CCTV merupakan alat yang efektif untuk melakukan monitoring tetapi untuk mengetahui hal-hal yang terjadi pada suatu waktu tentu harus memutar ulang rekaman sehingga memakan waktu yang lama.

Maka dari itu dilakukan penelitian ini untuk membuat sistem yang dapat mengenali aktivitas yang terekam atau tertangkap oleh kamera digital, sehingga pengguna dapat mengetahui aktivitas yang terjadi pada suatu video dengan mudah dan efisien. Adapun penelitian-penelitian serupa namun riset tersebut menggunakan metode yang berbeda, seperti beberapa riset lainnya pada [2] yang menggunakan metode Multilayer Perceptron (MLP) dan riset [3] yang menggunakan metode Convolutional Neural Network (CNN) .

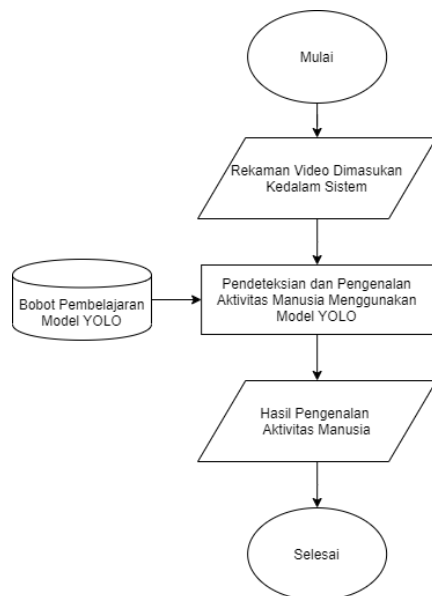
Dalam penelitian ini sistem akan dibuat dengan menggunakan metode yaitu Metode YOLO untuk pendeteksian dan pengenalan. Dengan menggunakan metode tersebut diharapkan sistem ini dapat mengenali suatu aktivitas yang terdapat dalam rekaman media perekam. Data yang akan diinput berupa citra foto ataupun video yang terdiri dari beberapa kelas aktivitas. Data aktivitas ini dipilih karena aktivitas ini merupakan aktivitas yang biasa dilakukan didalam sebuah ruangan belajar. Data tersebut berupa 5 jenis aktivitas aktivitas yaitu berdiri, duduk, belajar, mengangkat tangan, menepuk tangan.

## 2. Metode Penelitian

Sistem yang dibangun adalah program pengenalan aktivitas manusia dengan *You Only Look Once (YOLO)*. Sistem ini diharapkan dapat mengenali aktivitas yang terekam dalam video. Tahapan pada sistem secara umum dapat dilihat pada Gambar 1 dan Gambar 2.



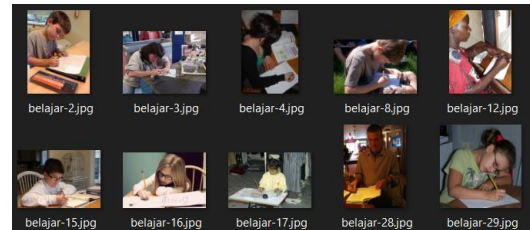
Gambar 1. Skema Pelatihan



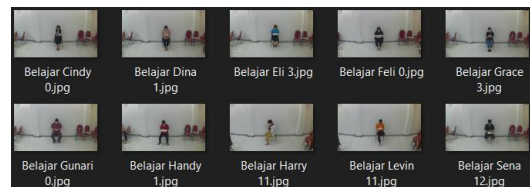
Gambar 2. Skema Pengujian

### 2.1 Dataset

Pada percobaan ini, dataset yang digunakan ada 2 yaitu dataset internet dan dataset IP Camera. Dataset internet merupakan dataset yang didapat dari internet yang berisi kumpulan data gambar-gambar aktivitas manusia. Dataset IP Camera merupakan dataset yang diambil sendiri dengan menggunakan IP Camera Xiaomi dengan tipe Home Camera. Contoh dataset citra internet dan IP Camera dapat dilihat pada Gambar 3 dan Gambar 4.



Gambar 3 Sample Dataset Internet



Gambar 4 Sample Dataset IP Camera

Setiap dataset terdiri dari kelas aktivitas manusia, seperti duduk, berdiri, belajar, mengangkat tangan, menepuk tangan. Dataset ini akan digunakan untuk pelatihan, validasi, dan pengujian. Tabel 1 menunjukkan jumlah dataset internet dan Tabel 2 menunjukkan jumlah dataset IP Camera.

Tabel 1 Dataset Yang Digunakan

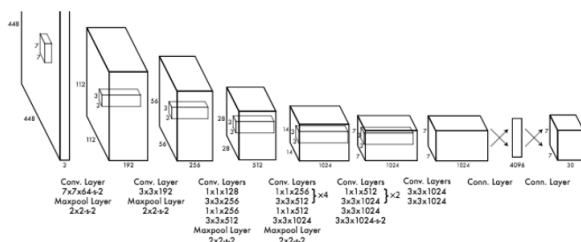
Jenis Aktivitas	Training		Testing	
	Internet	IP Camera	Internet	IP Camera
Berdiri	190	200	100	40
Duduk	190	200	100	40
Belajar	190	200	100	40
Mengangkat Tangan	190	200	100	40
Menepuk Tangan	190	200	100	40
Total	950	1000	500	200

### 2.2 You Only Look Once

YOLO merupakan sebuah pendekatan baru dalam sistem pendeteksian objek. YOLO melakukan pendeteksian dan pengenalan dengan sebuah jaringan syaraf tunggal (single neural network), yang memprediksi kotak-kotak pembatas dan probabilitas kelas secara langsung dalam satu evaluasi. Cara untuk

mendapatkan prediksi final yaitu dengan faktor penentunya adalah class confidence score. Class confidence score didapat berdasarkan probabilitas kondisional kelas dan box confidence score. Class confidence score merupakan cara mengukur nilai kepercayaan pada klasifikasi dan lokalisasi objek. Class confidence score memberikan nilai kepercayaan kelas spesifik kepada setiap kotak, yang mengkodekan kemungkinan terdapat kelas yang muncul di kotak dan seberapa sesuainya kotak yang diprediksi dengan objek atau kelas tersebut.

Struktur Jaringan YOLO seperti Convolution Neural Networks (CNN), tetapi yang membedakan YOLO hanya menggunakan convolutional layer dan pooling layer. Struktur Jaringan Yolo juga mirip seperti Google Net, yang membedakan yaitu YOLO menggunakan 1x1 convolutional layer (untuk integrasi antar channel) + 3x3 convolutional layer sebagai pengganti permulaan modul. Struktur pada YOLO v1 mengandung 24 convolutional layer dan 2 full connection layers [4]. Dapat dilihat pada Gambar 5.



### Gambar 5 Arsitektur YOLO

### 2.3 Konfigurasi Model Yolo

Nilai konfigurasi pada model Yolo yang digunakan berbeda-beda. Tabel 2 menunjukkan 4 model konfigurasi yang digunakan.

Tabel 2 Konfigurasi Model Yolo

Parameter	Konfigurasi Model			
	Model 1	Model 2	Model 3	Model 4
Batch	64	64	64	64
Subdivisions	32	32	16	16
Max_batches	4000	8000	8000	10000
Steps	3200, 3600	6400, 7200	6400, 7200	8000, 9000

## 2.4 Confusion Matrix

Metode confusion matrix adalah ringkasan hasil prediksi pada permasalahan klasifikasi. Metode ini digunakan untuk mengukur kinerja suatu metode klasifikasi. Metode ini merupakan hasil dari pengujian yang dibagi menjadi 4 kelas yaitu. True Positive (TP),

False Positive (FP), True Negative (TN) dan False Negative (FN), Keempat kelas tersebut dijelaskan sebagai berikut:

1. True Positive (TP) = hasil keluaran model yang benar dalam memprediksi kelas positive terhadap suatu data uji
2. True Negative (TN) = hasil keluaran model yang benar dalam memprediksi kelas negative terhadap suatu data uji.
3. False Positive (FP) = hasil keluaran model yang salah dalam memprediksi kelas positive terhadap suatu data uji.
4. False Negative (FN) = hasil keluaran model yang salah dalam memprediksi kelas negative terhadap suatu data uji.

Dengan menggunakan metode Confusion Matrix, Maka nilai tersebut dapat digunakan untuk menghitung nilai precision, recall, average precision (AP), F1-Score, Intersection over Union (IoU) dan mean Average Precision (mAP).

## 2.5 Precision

Nilai Presisi digunakan untuk memberi tahu jumlah data kategori positif yang diklasifikasikan dengan benar dibagi dengan total data yang diklasifikasi positif. Nilai presisi dapat dihitung dengan persamaan:

$$Presisi = \frac{TP}{TP - FP} \times 100\% \quad (1)$$

## 2.6 Recall

Nilai Recall adalah rasio dari total keseluruhan yang menunjukkan berapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem. Nilai Recall dihitung dengan persamaan sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

## 2.7 F1-Score

F1-Score bertujuan untuk menghitung kombinasi dari presisi dan recall. Nilai F1-Score didapatkan dari perbandingan rata-rata presisi dan recall yang dibobotkan. Nilai F1-Score dihitung dengan persamaan sebagai berikut:

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \times 100\% \quad (3)$$

## 2.8 Mean Average Precision (mAP)

Nilai Mean Average Precision adalah nilai rata-rata dari Average Precision (AP) yang membentuk matrik evaluasi guna mengukur kinerja dari pendeteksian. Nilai AP didapatkan dari perhitungan precision pada persamaan 1.

$$mAP = \sum_{i=1}^N \frac{AP(i)}{N} \times 100\% \quad (4)$$

## 2.9 Intersection over Union (IoU)

Nilai intersection over Union (IoU) adalah matriks evaluasi untuk mengukur keakuratan detektor pada objek. Intersection over Union (IoU) merupakan perbandingan antara ground-truth bounding box dengan prediksi bounding box seperti pada persamaan 5.

$$IoU = \frac{A \cap B}{A \cup B} \quad (5)$$

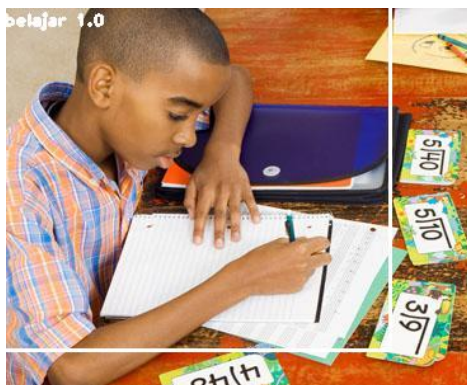
## 3. Hasil Percobaan

Dengan melakukan pelatihan menggunakan dataset latih menggunakan 4 model konfigurasi yang berbeda didapatkan nilai *confusion matrix*. Hasil dari performa model didapatkan nilai Mean Average Precision (mAP) dan IoU terbaik..

Dataset	Nilai mAP			
	Model 1	Model 2	Model 3	Model 4
Internet	87.60%	87.14%	88.28%	86.85%
IP Camera	99.96%	99.98%	99.90%	99.96%

Dataset	Nilai IoU			
	Model 1	Model 2	Model 3	Model 4
Internet	46.88%	46.48%	52.12%	50.67
IP Camera	86.78%	87.12%	87.89%	88.16%

Pengujian dilakukan dengan cara pengenalan aktivitas manusia. Model yang telah dilatih akan dilakukan pengujian untuk melakukan evaluasi hasil dari proses pelatihan. Contoh pengenalan aktivitas manusia dapat dilihat pada Gambar 6 dan Gambar 7.



Gambar 6 Contoh Hasil Pendeteksian Dataset Internet



Gambar 7 Contoh Hasil Pendeteksian Dataset IP Camera

Pengujian dilakukan dengan menggunakan dataset uji pada setiap kelas aktivitas. Pengujian dilakukan untuk melihat hasil akurasi terbaik akurasi yang akan didapatkan.

Dataset	Akurasi Model			
	Model 1	Model 2	Model 3	Model 4
Internet	73.4%	73%	74.6%	80.6%
IP Camera	97%	97%	96.5%	98%

Hasil Pengujian menunjukkan konfigurasi model yang berbeda maka mendapatkan hasil akurasi yang berbeda juga. Dapat bilang bahwa model yang telah dilatih cukup baik dalam mengenali aktivitas manusia.

## 4. Kesimpulan

Kesimpulan Yang dapat dari hasil pengujian yang dilakukan terhadap program Pendeteksian Aktivitas Manusia Menggunakan Metode You Only Look Once (YOLO) adalah sebagai berikut :

1. Program ini mampu melakukan pendeteksian aktivitas pada citra yang diinputkan dengan dataset internet cukup baik dengan hasil akurasi yang didapatkan yaitu 80.6% dengan dataset uji Internet. Serta mendapatkan nilai mAP 86.85% serta Average IOU 50.67% dengan dataset latih.
2. Program ini mampu melakukan pendeteksian aktivitas pada citra yang diinputkan dengan dataset IP Camera cukup baik dengan hasil akurasi yang didapatkan yaitu 98% dengan dataset uji Internet. Serta mendapatkan nilai mAP 99.96% serta Average IOU 88.16% dengan dataset latih.

Beberapa saran yang dapat diterapkan untuk pengembangan selanjutnya antara lain:

1. Mengganti dan menambahkan konfigurasi model yang berfungsi untuk memperjelas hasil deteksi aktivitas manusia.
2. Algoritma YOLO membutuhkan spesifikasi komputer yang cukup tinggi sehingga disarankan untuk penelitian selanjutnya menggunakan spesifikasi komputer yang lebih tinggi lagi sehingga pada proses pelatihan model dapat berjalan lebih cepat.
3. Menambahkan jumlah data latih untuk mendapatkan model yang lebih maksimal.

## **REFERENSI**

- [1] Eprints, Sejarah CCTV (Closed Circuit Television), <http://eprints.polsri.ac.id/1177/3/BAB%20II%20LA.pdf>.
- [2] Lina;Su, Jason, dan Ajenegoro, Daniel, "Human Activity Recognition Dari Rekaman Video Pengawas Dengan Metode Multilayer Perceptron", (Jurnal Muara Sains, Teknologi, Kedokteran, dan Ilmu Kesehatan, Vol. 5, No. 1, 2021).
- [3] Andrean, Jacob, Human Activity Recognition dari Rekaman Video Pengawas Menggunakan Convolutional Neural Network, Skripsi thesis, Universitas Tarumanagara 2021.
- [4] Aldhiyatika Amwin, Deteksi dan Klasifikasi Kendaraan Berbasis Algoritma You Only Look Once (YOLO), Skripsi thesis, Universitas Islam Indonesia, 2021.

**Harry Ronaldo Yudistira**, saat ini sebagai mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.

**Lina S.T., M.Kom., Ph.D.**, memperoleh gelar Sarjana dari Universitas Tarumanagara, Indonesia tahun 2001 dan gelar Magister dari Universitas Indonesia, Indonesia tahun 2004. Kemudian tahun 2009 memperoleh gelar Ph.D. dari Nagoya University, Jepang. Saat ini sebagai Dosen Tetap Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara