



**Pengembangan Metode Deteksi Gerakan Mencurigakan untuk  
Menjaga Keamanan**

**KUALIFIKASI**

**Yuditha Hatur Puspita**

**99223144**

**PROGRAM DOKTOR TEKNOLOGI INFORMASI**

**UNIVERSITAS GUNADARMA**

**Juni 2024**

## DAFTAR ISI

DAFTAR ISI.....	ii
PENDAHULUAN .....	1
1.1    Latar Belakang .....	1
1.2    Rumusan Masalah .....	3
1.3    Batasan Masalah Penelitian.....	4
1.4    Tujuan penelitian .....	4
1.5    Kontribusi dan Manfaat.....	4
BAB II.....	5
TELAAH PUSTAKA .....	5
2.1    Visi Komputer .....	5
2.2    Pengolahan Citra .....	5
2.3    Artificial Intelligence .....	6
2.4    Machine Learning.....	6
2.4.1    Supervised Learning.....	7
2.4.2    Unsupervised Learning .....	7
2.5    Deep Learning .....	7
2.6    Artificial Neural Network .....	8
2.7    Transformers .....	9
2.7.1    Vision Transformers.....	10
2.7.2    Detection Transformers.....	10
2.7.3    Patch Embedding .....	11
2.7.4    Layer Normalization .....	12
2.7.5    Multi-Head Attention .....	12
2.7.6    Multi Layer Perceptron .....	13
2.8    Confusion Matrix .....	13
2.9    Deteksi Objek .....	14
2.10    Robot .....	16
2.11    Bahasa Tubuh .....	19
2.12    Ekspresi Wajah.....	22
2.13    Alphapose .....	22

2.14	Python.....	24
2.15	Penelitian Terdahulu.....	24
BAB III .....		28
METODOLOGI PENELITIAN.....		28
3.1	Tahapan Penelitian .....	28
3.2	Akuisisi Data .....	29
3.3	Pre-processing .....	29
3.4	Augmentasi Data .....	30
3.5	Pelatihan Model.....	31
3.6	Evaluasi Model.....	31
3.7	Implementasi Model.....	32
3.8	Jadwal Penelitian.....	33
Bibliografi .....		34

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan teknologi pada penggunaan kecerdasan artifisial (*Artificial Intelligence*) menawarkan kemudahan dalam mendeteksi suatu objek dengan bantuan mesin. Salah satu teknologi yang dapat diterapkan adalah *Deep Learning*. *Deep Learning* merupakan subbidang *Artificial Intelligence* (AI) yang meniru kerja otak manusia dalam mengolah data dan menghasilkan pola untuk digunakan dalam pengambilan keputusan (Ashqar, B. A. M., Abu-Nasser, B. S., dan Abu-Naser, S. S., 2019). *Deep Learning* juga ampuh untuk ekstraksi fitur karena lebih unggul dalam memberikan informasi gambar yang lebih baik (Tan, J. W., Chang, S. W., Abdul-Kareem, S., Yap, H. J., dan Yong, K. T., 2018). Perkembangan deteksi objek saat ini semakin beranekaragam dan semakin populer, tidak hanya digunakan sebagai *image processing* tetapi juga digunakan untuk penglihatan dari robot. Pengaplikasian deteksi objek secara umum digunakan untuk mengenali benda-benda disekitar area dimana robot itu bergerak. Dibutuhkan banyak metode untuk sistem deteksi objek pada robot, terutama robot yang membutuhkan penglihatan yang lebih cepat dan akurat secara *realtime* (Chesher, C., Andreallo, F., 2021).

Inovasi dilakukan para peneliti untuk membuat robot yang semakin sempurna dan mendekati kemampuan yang dimiliki oleh manusia, salah satunya kemampuan untuk melihat dan mengenali objek. Pada robot pengindraan, untuk melihat secara jelas suatu objek maka digunakan kamera sebagai mata dari robot tersebut. Kamera merupakan perangkat keras yang memiliki fungsi untuk menangkap objek untuk menjadikan sebuah gambar atau merekam suatu bayangan potret pada lembaran film. Sistem pendeteksi objek pada robot beroda ini dapat diaplikasikan untuk menjangkau atau melihat pada situasi yang tidak dapat dijangkau oleh manusia atau tempat yang berbahaya yang dapat mengancam keselamatan (Chesher, C. *et al.*, 2021).

Perkembangan dunia robotika pada era revolusi industri 4.0 telah berkembang dengan pesat (Zhanybek, S., Shaikholla, S., Dikhanbayeva, D.,

Shehan, E., dan Turkeyilmaz, A., 2022). Salah satu perkembangan yang populer dilakukan saat ini adalah pembuatan sistem operasi pada robot. Dalam sektor industri terdapat 422.000 unit robot di seluruh dunia dan diperkirakan akan terus bertambah sebanyak 12% per-tahun hingga tahun 2022, pada tahun 2023 instalasi robot global diperkirakan akan tumbuh sebesar 7% menjadi lebih dari 590.000 unit dan pada tahun 2026 akan mencapai 700.000 unit (Müller, C., 2023). Saat ini, robot telah membantu manusia di banyak bidang diantaranya militer (Sanaullah, M., Akhtaruzzaman, M., 2022), industri (Sekhar, R., Shah, P., Iswanto, I., 2022) logistic (Ferreira, B., Reis, J., 2023), transportasi (Jin, Z., Li, D., Xiang, J., 2023), pertanian (Widianto, M. H., Juarto, B., 2023), medis (Alseddiqi, M., Najam, O., 2022) hingga keamanan (Gao, X., Zhang, Q., Li, M., Lan, B., Fu, X., dan Li, J., 2022).

Dalam era teknologi yang terus berkembang, peran robot dalam bidang keamanan menjadi semakin penting. Badan Pusat Statistik (BPS) dalam publikasi Statistik Kriminal 2023 mencatat bahwa terjadi 372.965 tindak kejahatan terjadi di Indonesia tahun 2022. Dalam hal ini berarti tingkat resiko terkena tindak kejahatan (crime rate) per 100 ribu penduduk mencapai 137 orang. Namun, tindak kejahatan bisa diminimalisir jika terdeteksi sedari dini, karena sebelum melakukan tindak kejahatan, seseorang biasanya akan memperlihatkan bahasa tubuh (body language), dengan bahasa tubuh, seseorang bisa memberikan petunjuk tentang niat atau kondisi emosional sebelum melakukan tindakan tertentu, termasuk niat kejahatan.

Robot telah menjadi bagian dalam upaya menjaga keamanan dan keselamatan, baik ditingkat individu maupun masyarakat secara luas. Kemampuan yang digunakan untuk melakukan tugas-tugas yang berpotensi berbahaya atau repetitif, serta mengakses area yang sulit dijangkau oleh manusia, robot membawa dampak yang signifikan dalam meningkatkan efektivitas sistem keamanan, baru-baru ini pengawasan keamanan telah menerapkan teknik pelacakan dan deteksi berbasis visual untuk meningkatkan kenyamanan dan keamanan bagi manusia (Chang, L. C., Pare, S., Meena, M. S., Jain, D., Li, D. L., Saxena, A., Prasad, M., dan Lin, C. T., 2020). Dalam konteks keamanan, robot dapat digunakan dalam berbagai peran, mulai dari pengawasan dan patroli hingga deteksi dan respons terhadap situasi darurat (Anbumani, P., Feloomi, K., Dharshana, B., dan Divya, S.,

2023). Robot dapat diprogram untuk melakukan pemantauan terus-menerus terhadap lingkungan, mengidentifikasi potensi ancaman, dan memberikan tanggapan cepat dalam situasi darurat. Robot juga dapat ditugaskan untuk tugas yang beresiko dan mengancam (Akilan, T., Chaudhary, S., Kumari, P., 2020). Selain itu, kemajuan dalam teknologi sensor, kecerdasan buatan, dan pemrosesan data memungkinkan robot untuk mengambil keputusan yang lebih cerdas dan beradaptasi dengan lingkungan mereka. Ini menghasilkan sistem keamanan yang lebih responsif, adaptif, dan efisien.

Penelitian ini mengusulkan sebuah sistem yang dapat mengurangi kejahatan secara preventif, tantangan yang saat ini dihadapi adalah perlu adanya model yang mampu memiliki kinerja dibawah pengaruh berbagai macam faktor, termasuk faktor ketidakstabilan lingkungan, pencahayaan yang berbeda dan sudut pandang yang dapat mencegah kejadian yang tidak diinginkan. Oleh karena itu, penelitian ini bertujuan untuk merancang sebuah model yang dapat mendeteksi ekspresi dan gerakan seseorang yang mencurigakan, menggunakan Vision Transformers kemudian dapat diimplementasikan ke Robot Operating System guna melakukan tindakan preventif sebelum kejahatan tersebut terjadi.

## **1.2 Rumusan Masalah**

Berdasarkan uraian pada latar belakang penelitian, maka dapat diidentifikasi rumusan masalah sebagai berikut:

1. Bagaimana membangun dataset sebagai data latih dan data uji untuk model deteksi gerakan yang mencurigakan?
2. Bagaimana mendeteksi gerakan mencurigakan menggunakan Vision Transformers?
3. Bagaimana mengembangkan model Vision Transformers untuk mendeteksi gerakan mencurigakan?
4. Bagaimana mengimplementasikan model Vision Transformers pada robot keamanan?

### **1.3 Batasan Masalah Penelitian**

Beberapa batasan dalam penelitian ini adalah:

1. Proses pendeteksian menggunakan Vision Transformers
2. Proses pengimplementasian menggunakan Robot Operating System
3. Data yang digunakan merupakan data primer yang berupa gerakan crossed arms, picking at/biting nails dan pacing sedangkan data sekunder digunakan untuk mendeteksi facial expression seperti neutral, happy, fear dan angry.

### **1.4 Tujuan penelitian**

Sehubungan dari rumusan masalah yang telah dipaparkan, tujuan dari penelitian ini sebagai berikut:

5. Membangun dataset sebagai data latih dan data uji untuk model deteksi gerakan yang mencurigakan.
6. Mendeteksi gerakan mencurigakan menggunakan Vision Transformers.
7. Mengembangkan model Vision Transformers untuk mendeteksi gerakan mencurigakan.
8. Mengimplementasikan model Vision Transformers pada robot keamanan.

### **1.5 Kontribusi dan Manfaat**

Dengan adanya penelitian ini diharapkan dapat memberikan manfaat dalam bidang ilmu, berupa model pengembangan metode pendeteksian ekspresi dan gerakan mencurigakan menggunakan Vision Transformers. Selain itu, penelitian ini juga berkontribusi dalam pengembangan robot untuk bidang keamanan, serta melindungi masyarakat dari tindak kejahatan yang tidak diharapkan.

## **BAB II**

### **TELAAH PUSTAKA**

#### **2.1 Visi Komputer**

Visi computer (computer vision) adalah bidang ilmiah interdisipliner yang membahas bagaimana computer dapat memperoleh pemahaman tingkat tinggi dari gambar atau video digital. Dari perspektif teknik, visi computer berusaha untuk memahami dan mengotomatisasi tugas-tugas yang dapat dilakukan oleh sistem visual manusia (Ballard et al., 1982). Sedangkan menurut (Matsuzaka et al., 2023) Visi computer (computer vision) adalah ilmu computer yang mempelajari bagaimana computer dapat “melihat”. Deep learning merupakan metode yang digunakan untuk mewujudkan visi computer melalui pengenalan gambar dan deteksi objek dengan cara memproses gambar yang diambil melalui kamera. Metode yang digunakan untuk memecahkan masalah dalam visi computer bergantung pada domain aplikasi dan sifat data yang dianalisis, visi computer sendiri memiliki berbagai jenis sub domain yang digunakan pada permasalahan tertentu, diantaranya seperti segmentasi gambar, deteksi objek, pengenalan wajah, deteksi tepi, deteksi pola, klasifikasi gambar dan pencocokan fitur.

#### **2.2 Pengolahan Citra**

Citra merupakan salah satu bentuk informasi yang diperlukan manusia selain teks, suara dan video. Informasi yang terkandung dalam sebuah citra dapat diinterpretasikan berbeda-beda oleh manusia satu dengan yang lain (Ratna, 2020).

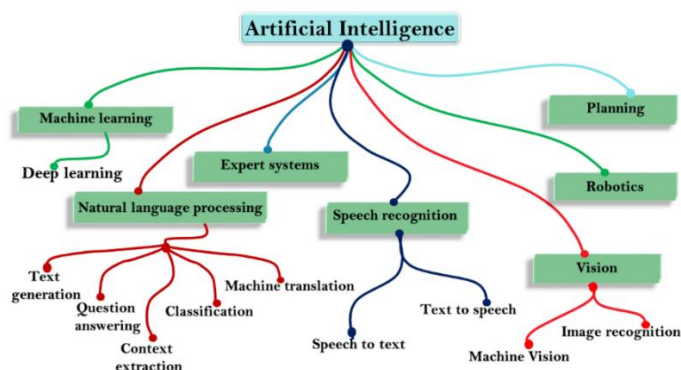
Pengolahan citra adalah suatu jenis teknologi untuk menyelesaikan masalah mengenai pemrosesan gambar, sedangkan *computer vision* mempunyai tugas untuk membuat suatu keputusan tentang objek fisik nyata yang didapat dari perangkat atau sensor. Pengolahan Citra Digital atau (*Digital Image Processing*) merupakan bidang ilmu yang mempelajari tentang bagaimana suatu citra itu dibentuk, diolah, dan dianalisis sehingga menghasilkan informasi yang dapat dipahami oleh manusia. Berdasarkan bentuk sinyal penyusunnya, citra dapat digolongkan menjadi dua jenis yaitu citra analog dan citra digital (Yuhandri et al., 2022).



Identifikasi objek pada citra digital membutuhkan teknik dan metode yang mampu untuk mengekstraksi dan mengidentifikasi fitur-fitur yang terdapat pada citra digital, dimana komponen utamanya adalah warna sebagai dasar dari representasi objek pada citra digital. Pengklasifikasi citra secara digital merupakan proses pengelompokan piksel-piksel kedalam kelas-kelas atau kategori-kategori yang telah ditentukan berdasarkan nilai kecerahan (digital number) piksel yang bersangkutan (Purwanto *et al.*, 2019).

## 2.3 Artificial Intelligence

*Artificial Intelligence* terdiri dari dua kata yaitu *Artificial* (buatan) dan *Intelligence* (kecerdasan), *artificial* berarti buatan manusia dan kecerdasan adalah kemampuan berfikir. Jadi dapat didefinisikan bahwa *Artificial Intelligence* adalah kecerdasan buatan dari cabang ilmu komputer yang dapat berperilaku, berfikir dan membuat keputusan seperti manusia sesuai dengan program yang telah dibuat (Bhbosale *et al.*, 2020). Beberapa penerapan yang dilakukan dalam *Artificial Intelligence* sebagai berikut ini.



Gambar 2.1. Penerapan *Artificial Intelligence* (Bhbosale *et al.*, 2020)

## 2.4 Machine Learning

*Machine Learning* merupakan bagian dari *Artificial Intelligence*, *Machine Learning* itu sendiri adalah metode yang menggunakan kumpulan data untuk meningkatkan kinerja (Sekrst *et al.*, 2023).

*Machine Learning* memungkinkan komputer untuk memiliki kemampuan mengidentifikasi pola, menemukan wawasan, memprediksi berdasarkan data dan

membuat keputusan berdasarkan data. Secara umum *machine learning* dapat dibagi menjadi dua tipe yang terdiri dari:

#### **2.4.1 Supervised Learning**

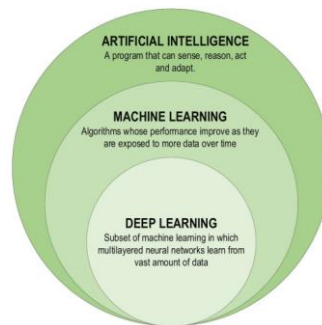
Tipe ini memiliki ciri dimana data yang digunakan oleh model merupakan *mapping* antara *input* dengan variabel target. Pada algoritma *Supervised Learning*, sistem diberikan *training data set* berupa informasi masukan dan keluaran yang diinginkan, sehingga sistem akan mempelajari berdasarkan data yang telah ada. Sistem akan mencari pola dari dataset, kemudian pola itu akan dijadikan sebagai acuan untuk kumpulan data berikutnya (Santoso, 2021).

#### **2.4.2 Unsupervised Learning**

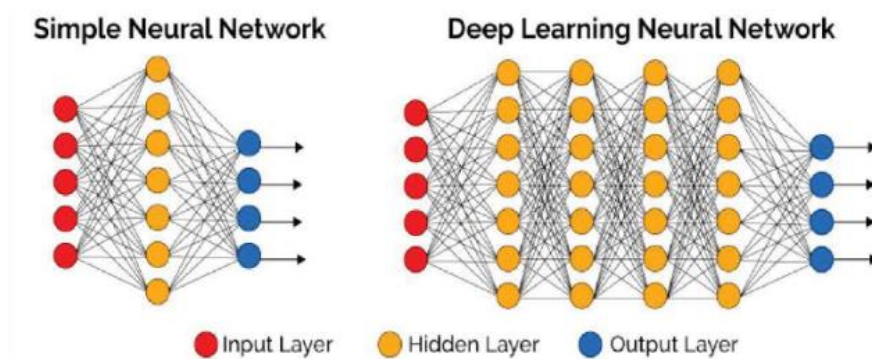
Tipe ini memiliki ciri dimana model pembelajaran dimaksudkan untuk mendeskripsikan atau mencari pola/hubungan pada data. Algoritma *Unsupervised Learning* bersifat deskriptif, yang akan berguna untuk mengelompokkan atau mengkategorikan data. Algoritma ini tidak mendapatkan *training data set*, karena algoritma ini bukan bersifat prediktif, sehingga membutuhkan pembelajaran dari data yang telah ada (Santoso, 2021).

### **2.5 Deep Learning**

*Deep Learning* merupakan cabang khusus dari *Machine Learning* yang memanfaatkan fungsional *Machine Learning* dan melampaui kemampuan *Machine Learning*. *Deep Learning* dapat diartikan sebagai struktur berlapis yang mampu mereplikasi struktur otak manusia. Keunggulan dari *Deep Learning* adalah memiliki performa yang terbaik dalam menyelesaikan masalah kompleks dan mampu beradaptasi serta memahami permasalahan baru dengan mudah.



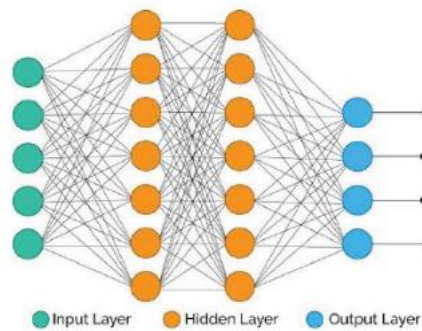
Gambar 2.2. AI, ML, dan DL (Alzubaidi *et al.*, 2021)



Gambar 2.3. Arsitektur *Deep Learning Neural Network* (Bre *et al.*, 2018)

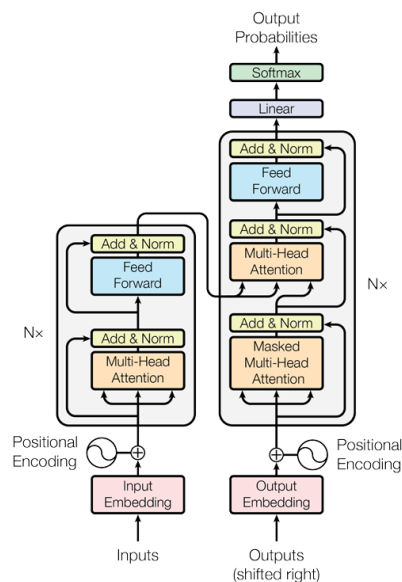
## 2.6 Artificial Neural Network

*Artificial Neural Network* atau Jaringan Syaraf Tiruan (JST) adalah model pemrosesan data berdasarkan cara sistem saraf biologis, seperti otak untuk memproses data. Jaringan Syaraf Tiruan (JST) berfokus pada struktur syaraf tetapi dengan skala yang lebih kecil. Banyak pakar kecerdasan buatan percaya bahwa Jaringan Syaraf Tiruan (JST) adalah yang terbaik dan harapan untuk merancang mesin yang cerdas (Dastres *et al.*, 2021). Jaringan saraf tiruan memiliki satu output yang digunakan sebagai input lain dan kemudian membuat jaringan. Berikut gambar dari arsitektur dari *Artificial Neural Network*.



Gambar 2.4. Arsitektur *Artificial Neural Network* (Bre *et al.*, 2018)

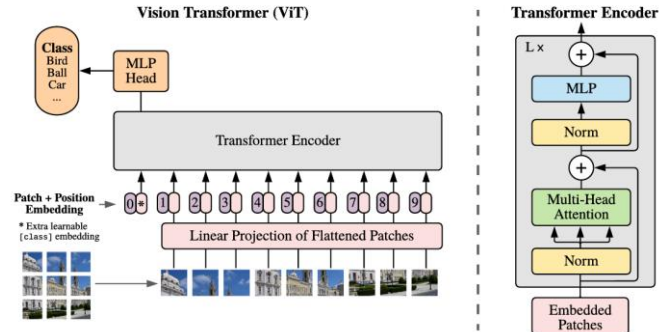
## 2.7 Transformers



Gambar 2.5 Arsitektur The Transformers (Vaswani *et al.*, 2017)

The transformer merupakan sebuah terobosan untuk Natural Language Processing, terutama untuk penggunaan penerjemah mesin yang bergantung pada attention mechanism, penelitian ini membangun arsitektur model baru yang bernama transformers, model ini menggunakan attention mechanism tanpa jaringan konvolusi untuk mencapai kualitas terjemahan yang lebih baik, namun model ini hanya dapat diterapkan dalam bentuk teks (Vaswani *et al.*, 2017).

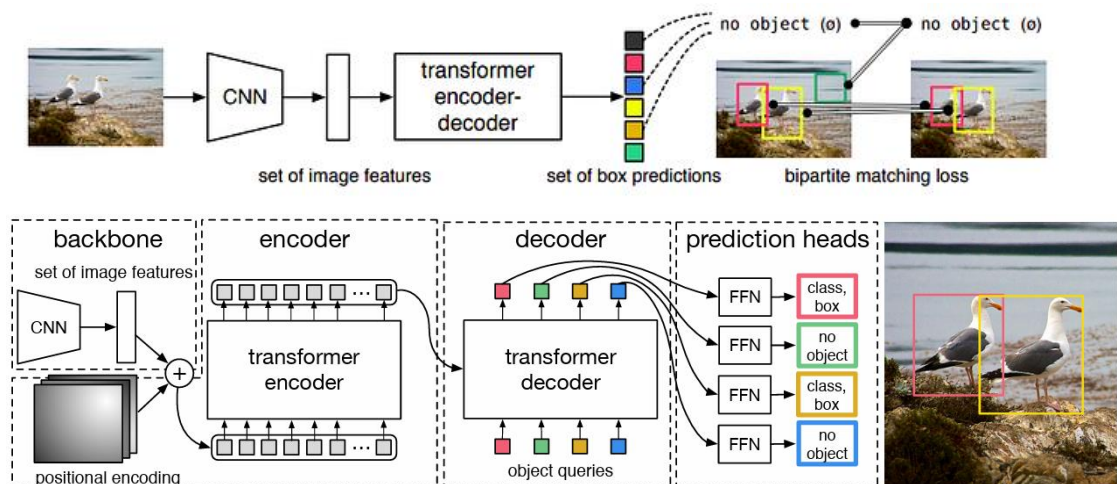
## 2.7.1 Vision Transformers



Gambar 2.6 Arsitektur Vision Transformer (Dosovitskiy et al., 2020)

Vision Transformer merupakan perkembangan dari model The Transformers yang hanya bisa diterapkan pada NLP, Vision Transformers dapat digunakan untuk memproses citra, menggunakan mekanisme self-attention untuk memproses patch gambar, yang memungkinkan model untuk mempelajari fitur yang berguna dari patch tanpa memerlukan lapisan konvolusional. Namun model ini tidak dapat diterapkan pada dataset yang berukuran kecil-sedang seperti ImageNet, karena tanpa regulasi yang kuat model ini akan menghasilkan akurasi dibawah CNN (Dosovitskiy et al., 2020).

## 2.7.2 Detection Transformers



Gambar 2.7 DETection Transformers (Carion et al., 2020)

Detection TRansformers atau dapat disingkat menjadi DETR merupakan perkembangan dari model Vision Transformers untuk deteksi objek, DETR dapat mendeteksi objek dengan memanfaatkan Convolutional Neural Network dengan model “transformer” dan menghilangkan komponen yang biasa dirancang manual seperti *non-maximum suppression* (NMS) dan *anchor*, kemudian menerapkan *Bipartite matching* dan *Hungarian Algorithm*.

Pada gambar 2.7 menjelaskan bahwa DETR menggunakan *backbone* dari CNN konvensional untuk mempelajari representasi 2D dari gambar input. Kemudian fitur tersebut diratakan (*flatten*) dan ditambah dengan positional encoding sebelum meneruskannya ke enkoder transformator. Dekoder transformator kemudian mengambil sebagai masukan sejumlah kecil penyematan posisi yang dipelajari, yang disebut kueri objek, dan sebagai tambahan memperhatikan keluaran enkoder. Kemudian meneruskan setiap penyematan keluaran dekode ke jaringan Feed Forward Network (FFN) yang memprediksi deteksi (kelas dan kotak pembatas) atau kelas “tanpa objek” (Carion et al., 2020).

### 2.7.3 Patch Embedding

Syarat data masukan untuk arsitektur vision transformer adalah citra tersebut dapat dibagi menjadi patch berukuran  $n \times n$ . Patch embeddings adalah proses dimana citra tersebut akan dibagi menjadi beberapa patch yang kemudian akan dijadikan vector satu dimensi berisi nilai piksel citra tersebut. Persamaan 1 akan menjelaskan rumus perubahan vector dua dimensi menjadi satu dimensi pada proses patch embeddings (Dosovitskiy et al., 2020).

$$X \in \mathbb{R}^{H \times W \times C} \rightarrow Xp \in \mathbb{R}^{N \times (P^2 \cdot C)} \quad (1)$$

$X$  = Citra masukan

$Xp$  = Citra masukan setelah transformasi

$H$  = Tinggi citra masukan

$W$  = Lebar citra masukan

$C$  = Jumlah channel citra masukan

P = Ukuran Patch

$$N = \frac{HW}{P^2}$$

#### 2.7.4 Layer Normalization

Proses pelatihan arsitektur deep learning pada umumnya menghabiskan sumber daya komputasi yang sangat banyak, sehingga dibutuhkan suatu cara yang dapat mengurangi penggunaan daya dan waktu yang dibutuhkan untuk menyelesaikan proses pelatihan tersebut. Salah satu langkah yang dapat ditempuh adalah Layer Normalization. Layer Normalization mengurangi penggunaan sumber daya dan waktu pelatihan dengan mengurangi nilai inputan menggunakan rata-rata dan standar deviasi. Semakin nilai inputan mendekati rata-rata dan standar deviasi, maka nilai inputan tersebut akan mendekati 0. Persamaan 2 berisi rumus menghitung layer normalization (Figo *et al.*, 2023).

$$x'_{i,k} = \frac{x_{i,k} - \mu^i}{\sqrt{\sigma_i^2 + \epsilon}} \quad (2)$$

Keterangan :

$x_{i,k}$  = Nilai inputan / Vektor

$\mu^i$  = Rata-rata nilai inputan / Vektor

$\sigma_i^2$  = Standar deviasi nilai inputan/ vektor

$\epsilon$  = Denominator

#### 2.7.5 Multi-Head Attention

Setelah Layer Normalization, Layer selanjutnya adalah Multi Head Attention. Di lapisan ini terjadi proses attention yang berfungsi untuk mencari informasi jangka Panjang yang berguna dari vektor masukan. multi head attention bekerja dengan cara menduplikasi vektor masukan menjadi query, key, dan value. Kemudian query, key, dan value tersebut akan diproses menggunakan scaled dot product. Proses tersebut akan dilakukan sesuai jumlah head dan kemudian hasilnya akan diconcat lalu dikompres menjadi ukuran vektor yang telah ditentukan. Gambar

3 akan mengilustrasikan proses yang terjadi selama proses scaled-dot product (Figo *et al.*, 2023).

### 2.7.6 Multi Layer Perceptron

Multi Layer Perceptron (MLP) adalah model jaringan syaraf tiruan yang terinspirasi dari system syaraf pada tubuh manusia. MLP terbagi menjadi 3 bagian yaitu lapisan masukan, lapisan tersembunyi, dan lapisan keluaran. Seperti system syaraf manusia, MLP memiliki neuron yang memiliki bobot. disetiap neuron tersebut ada fungsi aktivasi non-linear kecuali pada lapisan masukan (Figo *et al.*, 2023).

## 2.8 Confusion Matrix

*Confusion Matrix* merupakan sebuah tabel yang digunakan untuk melihat dan mengukur kinerja dari sebuah model klasifikasi. Tabel ini menampilkan dan membandingkan jumlah data yang telah diklasifikasi, *confusion matrix* akan menampilkan nilai aktual dengan nilai hasil prediksi model. Parameter yang dimiliki oleh *confusion matrix* ini diantaranya *accuracy*, *precision*, *recall* dan *f1-score*. Ada empat istilah yang mempresentasikan hasil klasifikasi pada *confusion matrix* yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN).

Tabel 2.1 *Confusion Matrix* (Ibnu Daqiqil Id, 2021)

		True Values	
		True	False
Prediction	True	TP Correct result	FP Unexpected result
	False	FN Missing result	TN Correct absence of result

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (2.6)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2.7)$$



$$Recall = \frac{TP}{(TP+FN)} \quad (2.8)$$

$$F-1 \text{ score} = \frac{(2*Recall*Precision)}{(Recall+Precision)} \quad (2.9)$$

Dimana:

TP = *True Positive* (Jumlah anggota kelas 1 berhasil diprediksi benar)

TN = *True Negative* (Jumlah anggota kelas 0 berhasil diprediksi benar)

FP = *False Positive* (Jumlah anggota kelas 1 gagal diprediksi benar)

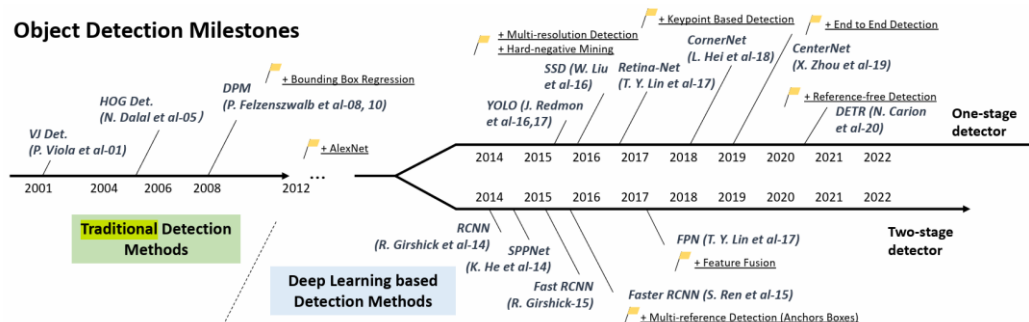
FN = *False Negative* (Jumlah anggota kelas 0 gagal diprediksi benar)

## 2.9 Deteksi Objek

Deteksi objek merupakan salah satu teknik dari bidang Computer Vision (Aningtiyas et al., 2020). Mengutip dari (Zou et al. 2023), deteksi objek adalah “an important computer vision task that deals with detecting instances of visual objects of a certain class (such as humans, animals, or cars) in digital images” atau bisa diartikan bahwa deteksi objek merupakan proses deteksi atau penemuan instances dari suatu objek visual kelas tertentu (seperti manusia, hewan, dan kendaraan) dalam sebuah citra digital. Deteksi objek yang merupakan masalah fundamental dari bidang Computer Vision memiliki tujuan untuk mengembangkan model dan teknik komputasi yang mampu menyediakan informasi paling dasar yang dibutuhkan oleh aplikasi Computer Vision yaitu: What objects are where? (Zou et al., 2013). Selain itu, deteksi objek juga meliputi berbagai tugas Computer Vision seperti segmentasi instance, Image Captioning (penamaan citra), serta Object Tracking (pelacakan objek). Deteksi objek banyak digunakan diberbagai bidang pekerjaan, mulai dari bidang kesehatan sampai ke bidang keamanan. Untuk bidang transportasi sendiri, deteksi objek biasanya digunakan untuk mendeteksi objek-objek yang dapat mengganggu kenyamanan berkendara seperti yang dilakukan oleh (Gao et al., 2018) pada penelitiannya mengenai klasifikasi objek pada lingkungan kendaraan autonomous.

Dalam dua dekade terakhir, deteksi objek telah melalui dua periode yaitu “periode deteksi objek tradisional (sebelum 2014)” dan “periode deteksi objek

berbasis Deep Learning (sesudah 2014)” (Zou et al., 2023). Berdasarkan Zou et al. (2023), “periode deteksi objek tradisional” adalah periode dimana deteksi objek dilakukan dengan menggunakan metode-metode yang terkesan tradisional karena pada saat itu belum ditemukan representasi citra yang efektif sehingga mengharuskan seseorang untuk mendesain representasi fitur yang rumit dan berbagai kemampuan yang mampu mempercepat proses komputasi untuk mengurangi penggunaan sumber daya yang pada saat itu tentunya relatif terbatas, sedangkan “periode deteksi objek berbasis Deep Learning” adalah periode dimana deteksi objek dilakukan dengan menggunakan metode-metode yang lebih efektif dan efisien karena model deteksi objek berbasis Deep Learning dapat mempelajari representasi fitur yang rumit dan bertingkat tinggi. Pada era deteksi objek berbasis Deep Learning, deteksi objek terbagi menjadi dua kategori yaitu “Two-stage Detection” dan “One-stage Detection” (Zou et al., 2023). Deteksi objek “Two-stage Detection” mempunyai dua tahapan dalam melakukan proses deteksi objek, tahap pertama yaitu memprediksi kandidat object proposal atau bounding box (kotak pembatas untuk mengidentifikasi posisi dan tipe objek) pada citra dan selanjutnya pada tahap kedua dilakukan proses klasifikasi dan regresi terhadap kandidat bounding box yang nantinya akan menghasilkan bounding box yang menunjukkan letak serta tipe objek, sebaliknya deteksi objek “One-stage Detection” hanya membutuhkan satu tahapan untuk melakukan proses deteksi objek (Lohia et al., 2021). Banyaknya tahapan yang dibutuhkan deteksi objek “Two-stage Detection” meningkatkan akurasi lokalisasi (penentuan posisi) dan rekognisi (pengenalan tipe) objek namun juga meningkatkan waktu yang dibutuhkan untuk melakukan proses deteksi, sebaliknya deteksi objek “One-stage Detection” mengurangi waktu yang dibutuhkan untuk melakukan proses deteksi namun juga mengurangi akurasi lokalisasi dan rekognisi objek (Zou et al., 2023). Tahapan umum pada deteksi objek “One-stage Detection” dan “Two-stage Detection”



Gambar 2.8 Roadmap perkembangan deteksi objek (Zou et al., 2023)

## 2.10 Robot

Robot adalah sebuah benda buatan manusia yang mampu melakukan pekerjaan yang dapat dilakukan oleh manusia dengan waktu yang jauh lebih singkat. Pada saat ini para ilmuwan sedang berusaha untuk menerapkan robot diberbagai bidang (Sachin *et al.*, 2021). Bidang tersebut diantaranya adalah bidang militer (Sanaullah *et al.*, 2022), bidang industri (Ravi Sekhar *et al.*, 2022), bidang logistic (Bárbara Ferreira *et al.*, 2023), bidang medis (Mohamed Alseddiqi *et al.*, 2022), bidang transportasi (Zibo Jin *et al.*, 2023), bidang pertanian (Mochammad Haldi Widiyanto *et al.*, 2023), hingga bidang keamanan (Xueshan Gao *et al.*, 2022). Hingga saat ini terdapat 5 jenis tipe robot yang sedang dikembangkan, yaitu:

### a. Pre-programmed Robots

Pre-programmed robots merupakan robot yang telah diprogram untuk melakukan satu tugas saja dan merupakan tugas sederhana. Contoh robot jenis ini adalah robot lengan mekanik yang digunakan untuk mengelas pintu.

### b. Humanoid Robots

Humanoid robots merupakan robot yang dapat melakukan kegiatan seperti manusia. Robot ini memiliki tampilan, perilaku dan suara yang mirip dengan manusia, robot ini dapat berlari, melompat, membawa benda dan sebagainya. Contoh robot jenis ini adalah robot Hanson Sophia dan Boston Dynamics Atlas, keduanya merupakan robot yang dapat melakukan pekerjaan manusia dengan mudah.

c. Autonomous Robots

Autonomous robots merupakan robot independen yang tidak membutuhkan intervensi manusia. Robot jenis ini mampu melakukan misi secara mandiri, misalnya robot penyedot debu.

d. Tele-operated Robots

Tele-operated robots merupakan robot yang dioperasikan atau dapat dikendalikan oleh manusia dari jarak jauh, robot ini bekerja di lingkungan dengan kondisi geografi yang ekstrim seperti cuaca buruk dan keadaan lainnya. Contoh dari tele-operated robots adalah kapal selam yang digunakan untuk mengatasi kebocoran pada saat terjadi tumpahan minyak atau drone yang digunakan untuk mendeteksi ranjau darat di medan perang.

e. Augmenting Robots

Augmenting robots merupakan robot yang dapat meningkatkan kapasitas manusia. Contoh dari robot jenis ini adalah eksoskeleton atau anggota tubuh palsu robotik.

### 2.10.1 Robot Operating System

Robot Operating System (ROS) merupakan sebuah kerangka kerja (framework) open-source yang digunakan untuk mengembangkan perangkat lunak untuk robot. Dikarenakan kompleksitas yang tinggi, framework ROS sangat diandalkan (Macenski et al., 2022). ROS mencakup *tools*, *libraries* dan *convention* yang bertujuan untuk menciptakan serta mengembangkan robot yang kompleks. Kerangka kerja ROS telah banyak digunakan selama lima belas tahun sejak diperkenalkan tahun 2009 (Quigley et al., 2009). Berikut merupakan daftar distribusi (distro) ROS yang dirilis hingga saat ini:

Tabel 2. Daftar Distro ROS 1

No	Nama Distro	Tanggal Rilis	Tanggal end-of-life
1	ROS Box Turtle	2 Maret 2010	-

2	ROS C Turtle	2 Agustus 2010	-
3	ROS Diamondback	2 Maret 2011	-
4	ROS Electric Emys	30 Agustus 2011	-
5	ROS Fuerte Turtle	23 April 2012	-
6	ROS Groovy Galapagos	31 Desember 2012	Juli 2014
7	ROS Hydro Medusa	4 September 2013	Mei 2015
8	ROS Indigo Igloo	22 Juli 2014	April 2019
9	ROS Jade Turtle	23 Mei 2015	Mei 2017
10	ROS Kinetic Kame	23 Mei 2016	April 2021
11	ROS Lunar Loggerhead	23 Mei 2017	Mei 2019
12	ROS Melodic Morenia	23 Mei 2018	Mei 2023
13	ROS Noetic Ninjemys	23 Mei 2020	Mei 2025

Sumber: <https://wiki.ros.org/Distributions>

Tabel 3. Daftar Distro ROS 2

No	Nama Distro	Tanggal Rilis	Tanggal end-of-life
1	Alpha1-alpha8	31 Agustus 2015	Desember 2016
2	Beta1	19 Desember 2016	Juli 2017
3	Beta2	5 Juli 2017	September 2017
4	Beta3	13 September 2017	Desember 2017
5	Ardent Apalone	8 Desember 2017	Desember 2018
6	Bouncy Bolson	2 Juli 2018	Juli 2019
7	Crystal Clemmys	14 Desember 2018	Desember 2019
8	Dashing Diademata	31 Mei 2019	May 2021
9	Eloquent Elusor	22 November 2019	November 2020
10	Foxy Fitzroy	5 Juni 2020	Juni 2023
11	Galactic Geochelone	23 Mei 2021	Desember 2022
12	Humble Hawksbill	23 Mei 2022	May 2027
13	Iron Irwini	23 Mei 2023	November 2024

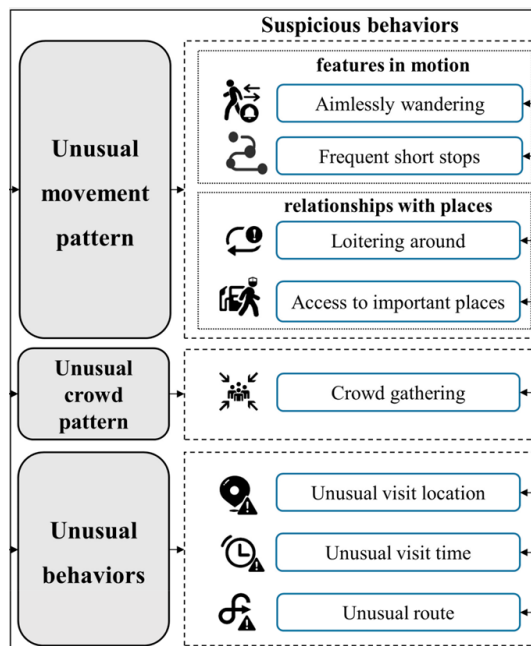
14	Jazzy Jalisco	Mei 2024	May 2029
----	---------------	----------	----------

Sumber: <https://docs.ros.org/en/rolling/Releases>

## 2.11 Bahasa Tubuh

Bahasa tubuh atau body language merupakan salah satu aspek penting dalam komunikasi nonverbal. Komunikasi nonverbal mencakup semua bentuk komunikasi tanpa kata-kata, dimana bahasa tubuh memainkan peran yang sangat signifikan. Bahasa tubuh meliputi gerakan tangan, kontak mata, postur tubuh, ekspresi wajah dan lain-lain. Pemahaman yang baik mengenai bahasa tubuh dapat meningkatkan efektivitas komunikasi dan membantu menghindari kesalahpahaman.

Bahasa tubuh juga dapat membantu para pakar dalam menganalisis sebuah kasus tindak kejahatan, sebelum melakukan tindak kejahatan, seseorang akan memperlihatkan bahasa tubuh yang mencurigakan dan tidak biasa. Pada gerakan mencurigakan merujuk pada perilaku atau aktivitas yang menimbulkan kecurigaan (Cheng et al., 2022).



Gambar 2.9 (Cheng et al., 2022)

Menurut (Cheng et al., 2022) gerakan mencurigakan meliputi; pola gerakan yang tidak biasa, pola kerumunan yang tidak biasa dan perilaku yang tidak biasa. Ciri dari pola gerakan yang tidak biasa antara lain mengembara atau berkeliling tanpa tujuan dan sering berhenti sejenak, sementara untuk pola kerumunan yang tidak biasa adalah berkumpulnya atau adanya kerumunan orang banyak, sedangkan untuk perilaku yang tidak biasa adalah lokasi kunjungan yang tidak biasa, waktu kunjungan yang tidak biasa dan rute yang tidak biasa dilalui.

Menurut artikel yang dimuat di healthline, gerakan gugup atau kegelisahan dapat dilihat dari gerakan kakinya, yaitu gerakan mengetuk-ngetuk kaki, kaki yang selalu bergoyang-goyang, dan menyilangkan kaki secara berpindah-pindah dari satu kaki ke kaki lainnya, sedangkan menurut Columbia University Medical Center, salah satu gerakan mencurigakan yang lain adalah menunjukkan sikap cemas atau gelisah, ada beberapa instrumen gerakan yang ditunjukkan oleh seseorang ketika merasa gelisah, namun tidak disadari. Tanda-tanda kegelisahan yang umum adalah sebagai berikut.

1. *Pacing*

*Pacing* merupakan gerakan gelisah yang paling umum dilakukan, yaitu berjalan kesana kemari dengan langkah yang konsisten dan berulang-ulang.

2. *Fidgeting*

*Fidgeting* merupakan gerakan kecil yang dilakukan terus-menerus dengan kaki, tangan, atau pun benda-benda disekitar dan sering kali dilakukan tanpa sadar. Seperti menggoyangkan kaki, memainkan pulpen atau meremas-remas tangan saat merasa cemas.

3. *Swaying*

*Swaying* merupakan gerakan bergoyang atau mengayunkan tubuh maju-mundur atau kesamping dengan ritme lambat dan berulang-ulang.

4. *Leaning*

*Leaning* merupakan gerakan menyender, biasanya akan memiringkan tubuh ke samping, depan ataupun belakang sebagai bentuk komunikasi non-verbal yang dapat mencerminkan ketidaknyamanan.

5. *Freezing*  
*Freezing* merupakan gerakan sebagai respons terhadap rasa takut, stress, kecemasan atau kejutan. Pada dasarnya tubuh tidak tahu apa yang harus dilakukan.
6. *Cracking Knuckles*  
*Cracking Knuckles* merupakan gerakan yang cukup umum dilakukan, yaitu menggerakkan jati-jari tangan, namun biasanya dilakukan sebagai kebiasaan atau karena stress.
7. *Crossed Arms*  
*Crossed Arms* merupakan gerakan menyilangkan tangan, gerakan ini dapat juga diartikan sebagai ketidaktertarikan, namun bisa juga dijadikan isyarat untuk menutup diri dan “melindungi” diri dari sekitar.
8. *Picking at or Biting Nails*  
*Picking at or Biting Nails* merupakan gerakan yang paling sering dilakukan, yaitu menggigit atau mengelupas kuku. Kegiatan ini merupakan tanda ketidaknyamanan yang sangat mencolok.
9. *Avoiding Eye Contact*  
*Avoiding Eye Contact* merupakan gerakan menghindari tatapan langsung dengan orang lain, jika seseorang tidak nyaman, kemungkinan besar mereka akan memalingkan muka dan memutus kontak mata.
10. *Voice Tells*  
*Voice Tells* merupakan perubahan dalam nada, kecepatan atau volume suara yang menunjukkan emosi atau keadaan pikiran seseorang. Suara bergetar merupakan indikasi kegelisahan.
11. *Blushing*  
*Blushing* merupakan reaksi alami tubuh yang tidak bisa dihindari, wajah akan memerah akibat peningkatan aliran darah ke kulit, biasanya disebabkan oleh rasa malu, gugup atau emosi yang kuat.



## 12. *Sweating*

*Sweating* merupakan respons tubuh yang mudah dikenali dan sulit untuk disembunyikan, sedikit berkeringat pada telapak tangan dan telapak kaki merupakan reaksi alami terhadap situasi gelisah.

### 2.12 Ekspresi Wajah

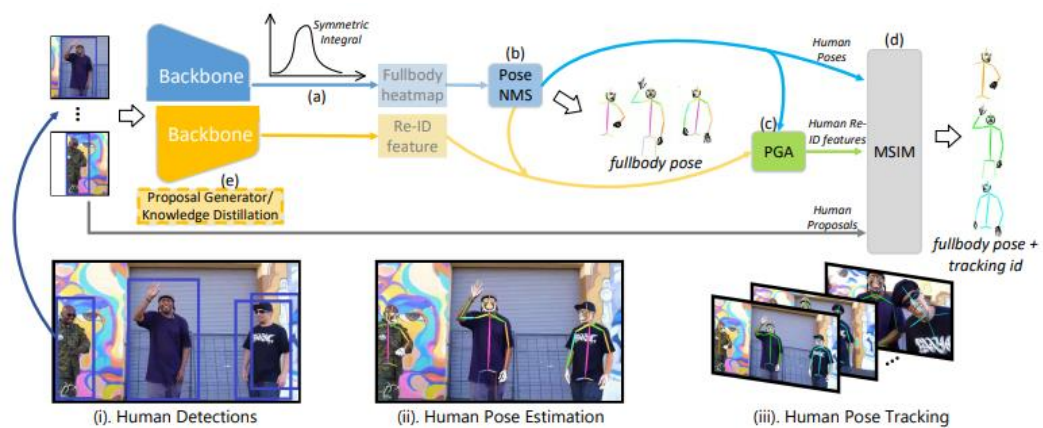
Ekspresi wajah adalah manifestasi visual dari emosi dan perasaan yang diekspresikan melalui perubahan posisi dan gerakan otot-otot wajah. Ekspresi wajah tidak hanya mencerminkan kondisi emosional individu, tetapi juga memainkan peran penting dalam komunikasi nonverbal antar manusia. Menurut Ekman (1972), terdapat enam emosi dasar yang diungkapkan melalui wajah yaitu kebahagiaan, kesedihan, ketakutan, kemarahan, kejijikan, dan kejutan. Ekspresi wajah juga dapat membantu psikologi klinis dalam mendiagnosis gangguan emosional dan mental seseorang, kemudian pada kasus orang gelisah, biasanya cenderung memberikan wajah ketakutan (Rosa, M. F., et al, 2017).

### 2.13 Alphapose

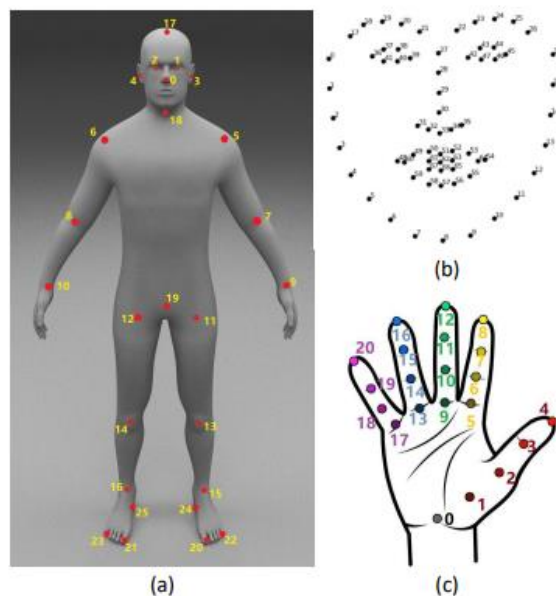
AlphaPose adalah teknologi deteksi pose tubuh manusia yang menggunakan metode deep learning untuk menghasilkan prediksi kerangka tubuh (skeleton) secara akurat dari gambar atau video. Teknologi ini mampu mendeteksi posisi sendi-sendi utama pada tubuh manusia, seperti bahu, siku, lutut, dan pergelangan tangan, dengan presisi tinggi. AlphaPose menggabungkan teknik deteksi objek dengan jaringan syaraf tiruan untuk menghasilkan peta pose tubuh yang terperinci. AlphaPose menggunakan beberapa komponen utama dalam arsitekturnya:

1. **Pose Estimation Network:** Jaringan ini dirancang untuk mendeteksi keypoints atau titik-titik kunci pada tubuh manusia dalam gambar. Ini melibatkan beberapa lapisan CNN yang dilatih dengan dataset pose tubuh.
2. **Part Affinity Fields (PAFs):** PAFs digunakan untuk menghubungkan titik-titik kunci dan membentuk kerangka tubuh. Ini memungkinkan algoritma untuk mempertimbangkan hubungan spasial antara sendi-sendi tubuh dan menghasilkan representasi pose yang lebih konsisten.

3. **Bottom-up vs. Top-down Approach:** AlphaPose mengadopsi pendekatan bottom-up yang dimulai dengan mendeteksi titik-titik kunci individu dan kemudian menghubungkannya menjadi pose lengkap. Ini berbeda dengan pendekatan top-down yang pertama-tama mendeteksi orang dalam gambar dan kemudian mendeteksi pose mereka.



Gambar 2.10 ilustrasi full body pose estimation dan tracking framework (Fang, H. S., 2022)



Gambar 2.11 Format keypoint untuk tubuh, wajah dan tangan (Fang, H. S., 2022)

## 2.14 Python

*Python* merupakan bahasa pemrograman *open-source* serbaguna yang dapat dijalankan di hampir semua arsitektur sistem, mulai dari *web*, *data science* hingga *machine learning*. *Python* memiliki Bahasa pemrograman yang dinamis dan diklaim mampu menggabungkan kapabilitas, kemampuan dengan sintaksis kode yang sangat jelas. Selain itu *Python* adalah Bahasa *cross-platform* yang mampu dijalankan di macam-macam OS seperti *windows*, *linux* serta *macOS*. *Python* memiliki banyak *library* yang berguna untuk *data analysis* dan *machine learning*, diantaranya adalah *pandas*, *matplotlib*, *numpy*, *scipy*, *tensorflow* dan lain-lain.

## 2.15 Penelitian Terdahulu

Penelitian dalam pendeteksi gerakan mencurigakan telah dilakukan oleh peneliti sebelumnya, berikut merupakan beberapa penelitian dalam pendeteksi gerakan mencurigakan dan ekspresi pada wajah.

Lee, J., Lee, S., Cho, W., Siddiqui, Z. A., dan Park, U. (2021) mengusulkan pengawasan berbasis CCTV menggunakan arsitektur Modified Time-Series Vision Transformer (TSViT) yang dapat mendeteksi gerakan mencurigakan yaitu mengikuti seseorang (*tailing*), penggunaan CCTV ini dapat mendeteksi *anomaly*, tujuan dari mendeteksi *anomaly* tersebut adalah untuk menemukan kelainan dalam pola berjalan pejalan kaki. Hasil eksperimental menunjukkan bahwa TSViT dapat melakukan deteksi objek lebih baik daripada metode CNN. TSViT dapat digunakan dalam banyak aplikasi untuk deteksi anomali video, bahkan dengan dataset kecil.

Singh, S., Dewangan, S., Krishna, G. S., Tyagi, V., Reddy, S., Medi, P. R. (2022) mengusulkan pengawasan berbasis CCTV menggunakan arsitektur Video Vision Transformers (ViViT) untuk mendeteksi kekerasan, arsitektur ini dapat mengestimasi keadaan kekerasan yang efisien dalam klip video. Model ini digunakan untuk mengkategorikan sebuah tindakan termasuk kekerasan atau non kekerasan. Namun, arsitektur ini membutuhkan sejumlah data yang besar agar hasil deteksi dapat menjanjikan.

Lee, M. R., dan Shih, Z. (2022) mengusulkan robot pengawasan yang menggunakan arsitektur Convolutional Neural Network (CNN), arsitektur ini

menggunakan visual perception dan control task, untuk visual perception berfungsi untuk mendeteksi gerakan, modul pelacakan, mendeteksi wajah dan modul pengenalan serta control task yang berfungsi untuk control gerak dan navigasi pada robot. Tiga model yang dilakukan pada penelitian ini adalah model CNN yakni AlexNet, VGGNet dan GoogleNet. VGGNet memiliki kinerja terbaik dibawah pengaruh berbagai macam faktor, namun model ini memiliki jumlah parameter terbesar dan eksekusi yang paling lambat. Dari ketiga model, tidak satu pun model yang dapat mencapai pengenalan wajah ketika dibawah oklusi atau ketika gaya rambut target berubah secara signifikan. Meskipun algoritma pengenalan wajah yang diusulkan tidak dapat mencapai pengenalan wajah di bawah semua faktor yang tidak stabil, kombinasi algoritma deteksi wajah frontal dan 2D CNN modelis dalam skala kecil dan relatif cepat dalam eksekusi. Disarankan untuk mengganti algoritma deteksi objeknya ke algoritma yang lebih kompleks.

Shreyash C., Tiwari, R. N., Siddique, S., Jain, P., dan Mane, S. (2023) penelitian ini mengusulkan arsitektur Convolutional Neural Network (CNN) yakni VGG-16 dan Long-term memory models (LSTM). CNN digunakan untuk mengekstrak fitur dari gambar agar mengurangi kompleksitas input, sedangkan LSTM digunakan untuk memproses aliran video untuk tujuan klasifikasi. Penelitian ini menghasilkan akurasi sebesar 87,15%. Tujuan dari penelitian ini adalah menggunakan video CCTV untuk kegiatan yang mencurigakan di area kampus, nantinya sistem akan memberi tahu otoritas yang berwenang melalui SMS ketika mendeteksi perilaku yang mencurigakan.

(Meddeb, H., Abdellaoui, Z., dan Houaidi, F., 2023) penelitian ini mengusulkan robot IoT sebagai robot pengawasan keamanan menggunakan algoritma Haar Cascade Classifier dan Local Binary Pattern Histogram (LBPH) dengan Sensor Infrared Pasif (PIR). Dalam penelitian ini robot IoT akan mendeteksi wajah secara real-time untuk memeriksa wajah yang dikenal dan wajah yang tidak dikenal, wajah yang tidak dikenal akan diidentifikasi sebagai penyusup. Namun robot IoT ini memiliki jangkauan yang terbatas.

(He, F., Zhang, L., 2023) penelitian ini mengusulkan robot keamanan didalam ruangan menggunakan Robot Operating System (ROS), Adaptive Monte Carlo

Localization (AMCL) package, Simultaneous Localization and Mapping (SLAM) dan You Only Look Once (YOLOv3) dengan tingkat akurasi sebesar 78,6%. Pada penelitian ini diharapkan untuk fungsi penentuan posisi, navigasi, dan deteksi target dapat digabungkan, sehingga robot dapat menyelesaikan tugas deteksi target dalam proses navigasi dan mengoptimalkan struktur sistem dan koordinasi serta kerja sama antara kedua fungsi, dan meningkatkan kinerja algoritma navigasi dan algoritma deteksi objek.

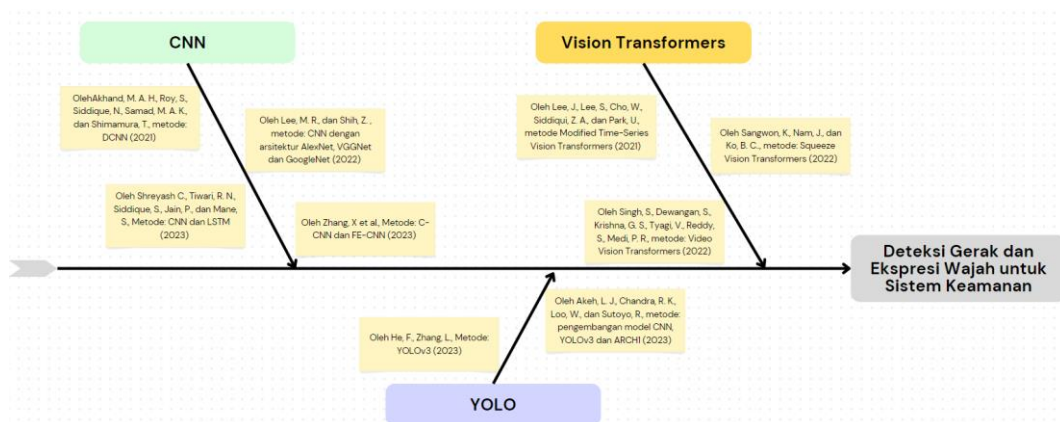
Sedangkan untuk penelitian terkait facial emotion (Akhand, M. A. H., Roy, S., Siddique, N., Samad, M. A. K., dan Shimamura, T., 2021) melakukan penelitian pengembangan metode facial emotion recognition (FER) yang efisien menggunakan model DCNN yang menangani tantangan melalui transfer learning, investigasi model dengan delapan model DCNN pra-terlatih populer pada gambar wajah bench mark dengan tampilan depan dan tampilan profil (dimana hanya satu mata, telinga, dan satu sisi wajah yang terlihat). Penelitian ini mengenalkan strategi pelatihan pipeline untuk penyempurnaan model secara bertahap hingga akurasi pengenalan yang tinggi. Namun, kumpulan data dengan gambar resolusi rendah atau dengan kasus yang sangat tidak seimbang akan memerlukan pra pemrosesan tambahan dan modifikasi yang sesuai dalam metode ini.

(Sangwon, K., Nam, J., dan Ko, B. C., 2022) mengusulkan squeeze ViT, sebuah metode untuk mengurangi kompleksitas komputasi dengan mengurangi jumlah dimensi fitur sekaligus meningkatkan kinerja FER dengan menggabungkan fitur global dan lokal secara bersamaan. Untuk mengukur kinerja FER dari Squeeze ViT, percobaan dilakukan pada dataset FER yang dikendalikan laboratorium dan dataset FER umum. Melalui eksperimen komparatif dengan pendekatan mutakhir sebelumnya, kami membuktikan bahwa metode yang diusulkan mencapai kinerja yang sangat baik pada kedua jenis kumpulan data. Penelitian ini fokus pada pengembangan algoritma FER berbasis ViT untuk memungkinkan FER yang akurat ketika menerapkan berbagai dataset.

(Akeh, L. J., Chandra, R. K., Loo, W., dan Sutoyo, R., 2023) penelitian ini mengembangkan model yang berhasil mengenali emosi dengan akurasi 98,9%, lebih baik daripada model CNN, YOLOv3 dan ARCH1, tidak menggunakan

augmentasi data dan tidak menggunakan dropout layer untuk memperoleh hasil terbaik. Namun, untuk ekspresi loss, fear and disgust menjadi yang paling sulit dikenali dan dataset masih belum seimbang.

(Zhang, X et al., 2023) melakukan penelitian tentang deteksi wajah perbedaan objektif facial expressions antara pasien penderita schizophrenia dan tidak menderita schizophrenia (sehat) menggunakan C-CNN dan FE-CNN dengan menekankan penggunaan strategi augmentasi data untuk meningkatkan kualitas data.



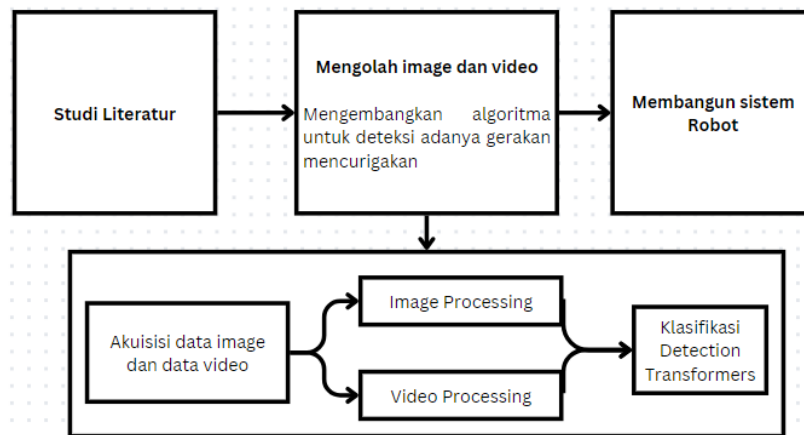
Gambar 2.10. Fishbone

## BAB III

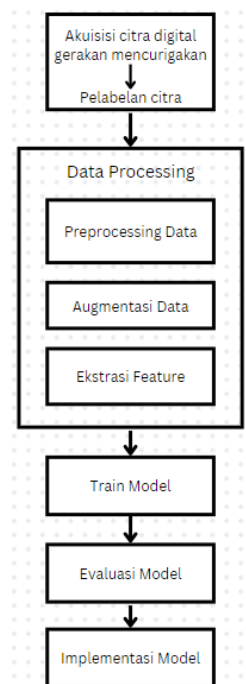
### METODOLOGI PENELITIAN

#### 3.1 Tahapan Penelitian

Penelitian ini dilakukan dalam beberapa tahapan. Tahapan yang dilakukan dimulai dari akuisisi data, preprocessing data, augmentasi data, melatih model, mengevaluasi model dan mengimplementasikan model.



Gambar 3.1 Tahapan Penelitian



Gambar 3.2 Tahapan Pengembangan Sistem

### 3.2 Akuisisi Data

Tahap pertama yang dilakukan adalah akuisisi data atau pengumpulan data, data yang digunakan dalam penelitian ini adalah data primer yang dikumpulkan oleh peneliti dan data sekunder yang diperoleh dari kaggle. Pada saat proses pengumpulan data primer, peneliti akan merekam 4 gerakan mencurigakan yakni menyilangkan tangan, menggigiti kuku, mengetuk-ngetuk kaki dan kaki bergoyang-goyang menggunakan ponsel selama 30 detik. Kemudian data akan dianotasi, sehingga setiap gerakan akan diberikan label.

Untuk data sekunder, akan digunakan dataset yang diperoleh dari kaggle dengan nama KDEF dataset, dataset ini dikembangkan oleh Karolinska Institute, Departemen Ilmu Saraf bagian Psikologi. Dataset tersebut berisi 4900 gambar dari 70 individu dengan tujuh keadaan emosional dari 5 sudut pandang berbeda. Pada penelitian ini hanya menggunakan lima keadaan emosional yaitu bahagia (happy), marah (angry), sedih (sad), takut (afraid) dan netral (neutral).



Gambar 3.3 Sample images dari dataset KDEF

### 3.3 Pre-processing

Sebelum data digunakan, preprocessing adalah tahap pertama dalam pengolahan data, kumpulan data yang telah kumpulkan atau didapatkan akan melalui tahapan preprocessing dengan tujuan agar dataset siap digunakan oleh model untuk dipelajari. Salah satunya pada data yang akan digunakan pada deteksi gerakan, data berupa video diubah menjadi beberapa frame dengan jumlah tetap. Untuk menghindari komputasi yang intensif, maka frame akan diperkecil

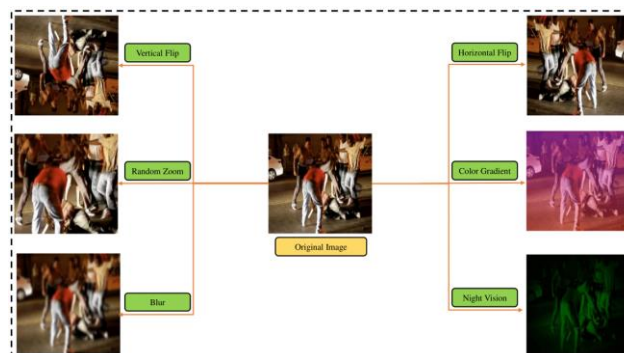


resolusinya. Namun, saat resolusi diperkecil rasio aspek harus tetap sama, karena perubahan rasio aspek dapat menyebabkan hilangnya beberapa informasi penting dalam video. Pekerjaan ini dilakukan dengan premis bahwa 56 frame berurutan cukup untuk mengilustrasikan peristiwa yang mencurigakan, karena biasanya kecepatan frame yang biasa dihasilkan oleh klip video adalah sekitar 25-30 frame per detik. Proses akan diolah menggunakan Alphapose dan dilakukan ekstraksi nilai keypoint. Kemudian koordinat keypoint akan dinormalisasi untuk setiap frame agar sesuai dengan skala yang seragam.

Sedangkan untuk dataset yang mendeteksi ekspresi akan diubah ukurannya menjadi ukuran piksel tetap yang lebih kecil dengan tetap mempertimbangkan konsep rasio aspek yang sama antara height dan width channels. Setiap frame akan diubah menjadi 224x224pixel dan displit menjadi 16x16 size patch.

### 3.4 Augmentasi Data

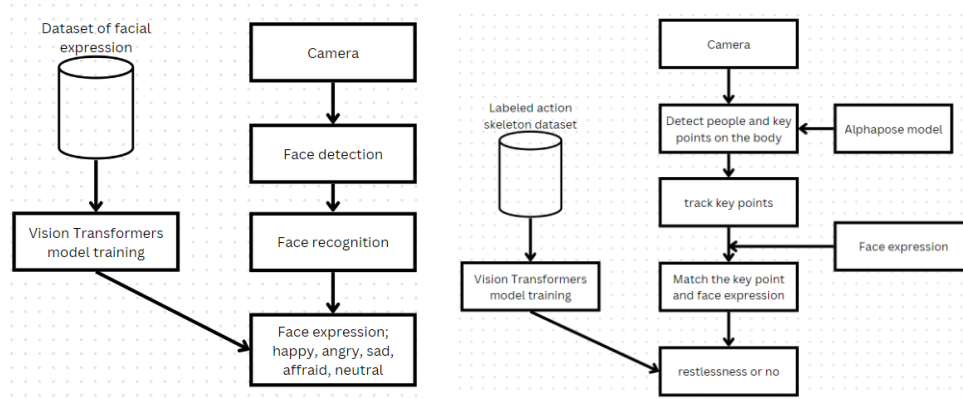
Karena Vision Transformers (ViT) tidak memiliki jaringan konvolusi, kinerja ViT lebih efektif jika dilatih sebelumnya pada dataset berskala besar. Maka pada penelitian ini menggunakan teknik augmentasi data untuk memvariasi data. Dengan demikian, data gambar yang telah diproses sebelumnya kemudian ditransformasikan dengan beberapa teknik augmentasi gambar seperti random flip, random zoom, blur, color gradient dan night version. Sedangkan untuk data gerakan, dilakukan transformasi warna seperti kecerahan, kontras dan saturasi pada frame.



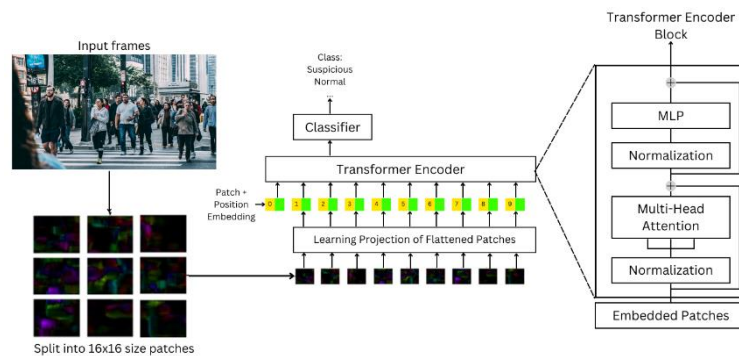
Gambar 3.4 Single Frame Augmentation

### 3.5 Pelatihan Model

Setelah data selesai dipreprocessing dan diaugmentasi, kemudian data dibagi menjadi dua kelompok yakni data latih dan data uji dengan komposisi 8:2. Alur pelatihan model ditunjukkan pada gambar 3.5 dan arsitektur vision transformers ditunjukkan pada gambar 3.6



Gambar 3.5(a) Face recognition using vision transformers 3.5 (b) Motion detection using Alphapose dan Vision Transformers



Gambar 3.6 Arsitektur Vision Transformers

### 3.6 Evaluasi Model

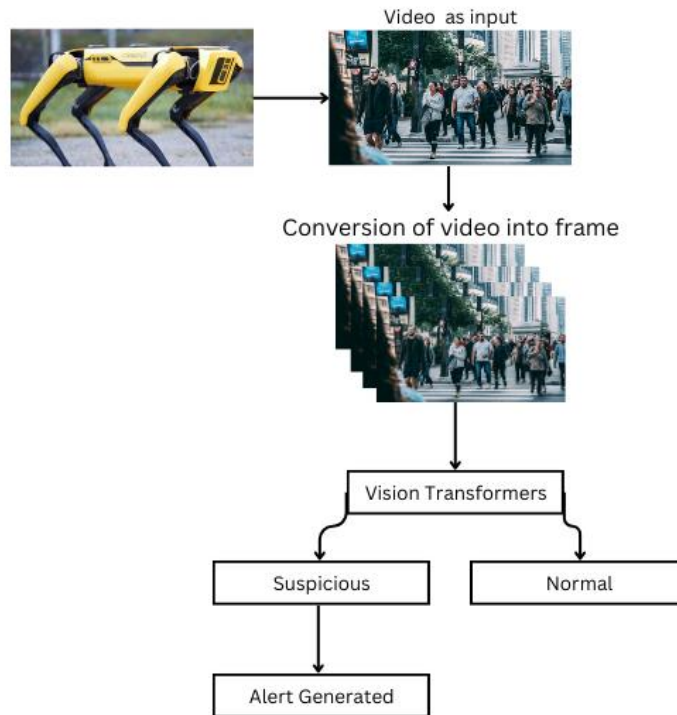
Tahap terakhir yang dilakukan terhadap model adalah tahap evaluasi, pada tahap ini dilakukan untuk mengukur kinerja model terhadap data testing. Penelitian ini menggunakan confusion matrix untuk menganalisis performa model klasifikasi. Confusion matrix adalah tabel yang digunakan untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi model dengan label sebenarnya pada data

uji atau data yang belum pernah dilihat oleh model. Dengan menggunakan nilai-nilai pada confusion matrix maka dapat ditentukan nilai evaluasi model, seperti accuracy, precision, recall dan f1-score. Nilai accuracy dapat menunjukkan seberapa akurat model dapat melakukan pendeteksian dengan benar terhadap keseluruhan data. Nilai precision adalah nilai yang menunjukkan ketepatan model dalam memprediksi sebuah class dengan benar. Nilai recall yang tinggi menunjukkan model dapat mengenali class dengan baik. Sedangkan, f1-score merupakan Harmonic Mean antara precision dan recall. Nilai f1-score yang tinggi mengindikasikan model memiliki precision dan recall yang baik.

### **3.7 Implementasi Model**

Setelah model berhasil dibuat, tahap terakhir dalam metodologi ini adalah menghubungkan model pendeteksi gerakan mencurigakan pada kamera yang berfungsi sebagai mata robot, model akan diimplementasikan pada robot keamanan untuk melihat apakah model vision transformer yang terlatih dapat digunakan untuk robot keamanan.

Pertama-tama membuat workspace pada Robot Operating System (ROS) yang baru, karena pelatihan model vision transformers telah selesai dilakukan pada tahapan sebelumnya, maka hanya perlu menyalin file tersebut kedalam folder. Sebelum melakukan pendeteksian, instal package `usb_cam` terlebih dahulu, package tersebut digunakan untuk membaca gambar atau video melalui camera. Setelah package driver camera telah diunduh, image topic dapat dapat dirilis dan dapat diamati pada virtual machine PC. Pada saat ini, file peluncuran yang sesuai dengan paket fungsi dapat dieksekusi untuk deteksi objek.



Gambar 3.7 Alur sistem robot

### 3.8 Jadwal Penelitian

No	Kegiatan	Tahun 1		Tahun 2		Tahun 3	
		Sem1	Sem2	Sem1	Sem2	Sem1	Sem2
1.	Kualifikasi						
2.	Persiapan penelitian						
3.	Penelitian						
4.	Progress report 1						
5.	Progress report 2						
6.	Sidang Terbuka						
7.	Sidang Tertutup						

## Bibliografi

- Alseddiqi, M., AlMannaei, B., Najam, O., A. Al-Mofleh, A. (2022). The Importance of Medical Robots in Improving Healthcare Services. *International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, pp. 1397-1401, <https://doi.org/10.1109/ICACRS55517.2022.10029105>
- Ashqar, B. A. M., Bassem, S. A. N., Samy S. A. N. (2019). Plant Seedlings Classification Using Deep Learning, *International Journal of Academic Information Systems Research (IJAISR)*, 3(1), 7-14.
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., (2020). End-to-End Object Detection with Transformers. In Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I. Springer-Verlag, Berlin, Heidelberg, 213–229. [https://doi.org/10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13)
- Chang, L. C., Pare, S., Meena, M. S., Jain, D., Li, D. L., Saxena, A., Prasad, M., Lin, C. T., (2020). An Intelligent Automatic Human Detection and Tracking System Based on Weighted Resampling Particle Filtering. *Big Data and Cognitive Computing*. 4(4), 27. <https://doi.org/10.3390/bdcc4040027>
- Chesher, C., Andreallo, F. (2022). Eye Machines: Robot Eye, Vision and Gaze. *Int J of Soc Robotics* **14**, 2071–2081. <https://doi.org/10.1007/s12369-021-00777-7>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Hounsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv, abs/2010.11929*.

- Ferreira, B., & Reis, J. (2023). A Systematic Literature Review on the Application of Automation in Logistics. *Logistics*, 7(40).  
<https://doi.org/10.3390/logistics7040080>
- Gao, X., Zhang, Q., Li, M. *et al.* (2022). Mobile Robot Combination Autonomous Behavior Strategy to Inspect Hazardous Gases in Relatively Narrow Man–Machine Environment. *Chinese Journal of Mechanical Engineering*.  
<https://doi.org/10.1186/s10033-022-00798-x>
- Iqbal, Z., M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Computers and Electronics in Agriculture*, 153:12–32
- Jin, Z., Li, D., Xiang, J. (2023). Robot Pilot: A New Autonomous System Toward Flying Manned Aerial Vehicles. *Engineering*, 27(8), 242-253.
- K. Trang, L. Tonthat, N. Gia Minh Thao, and N. Tran Ta Thi. (2019). Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning. *IEEE Conf. Sustain. Util. Dev. Eng. Technol. CSUDET 2019*, 138–142.
- Rajbongshi, A., Thaharim K., Md. Mahbubur, R., Anik, P., Shah, Md, T, S., Narayan, R., C. (2021). Recognition of mango leaf disease using convolutional neural network models: a transfer learning. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(3):1681-1688.
- Suleiman, Z., Shaikholla, S., Dikhanbayeva, D., Shehab, E., Turkyilmaz, A., & Zhou, Z. (2022). Industry 4.0: Clustering of concepts and characteristics. *Cogent Engineering*, 9(1).  
<https://doi.org/10.1080/23311916.2022.2034264>

- Tan, J., W., et al. (2018). Deep Learning for Plant Species Classification using Leaf Vein Morphometric, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(1), 82-90.
- Uparkar, O., Bharti, J., Pateriya, R. K., Gupta, R., (2023). Vision Transformer Outperforms Deep Convolutional Neural Network-based Model in Classifying X-ray Images, *Procedia Computer Science*, 218(10), 2338-2349.
- Wundt, W., Principles of physiological psychology, 1873, in: Readings in the history of psychology, East Norwalk, CT, US: Appleton-Century-Crofts., 1948 pp. 248–250
- Li, W., Passama, R., Bonnet, V., and Cherubini, A., (2023). "A comparison of human skeleton extractors for real-time human-robot interaction," *IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO)*, Berlin, Germany, 2023, pp. 159-165, <https://doi.org/10.1109/ARSO56563.2023.10187411>
- Lee J, Lee S, Cho W, Siddiqui ZA, Park U., (2021). Vision Transformer-Based Tailing Detection in Videos. *Applied Sciences*. 11(24):11591. <https://doi.org/10.3390/app112411591>
- Lee M-FR, Shih Z-S., (2022). Autonomous Surveillance for an Indoor Security Robot. *Processes*. 10(11),2175. <https://doi.org/10.3390/pr10112175>
- Chloe, S., Tiwari, R. N., Siddique, S., Jain, P., Mane, S., (2023). Detecting Suspicious Activities In Surveillance Videos Using Deep Learning Methods. *International Research Journal of Modernization in Engineering Technology and Science*. 5(1), 1804-1810. <https://www.doi.org/10.56726/IRJMETS33208>