Deteksi Objek & Klasifikasi Citra berbasis Deep Learning

Fauzan Firdaus

Introduction

Bio

Fauzan Firdaus (AI Instructor – Orbit Future Academy)
Sumedang, 10 Agustus 1998

Educations

S1 Informatika – Univ. Telkom (2016-2020)
S2 Informatika – ITB (2020-2022)

Contact

faoezanfirdaoes@gmail.com
linkedin.com/in/faoezanf

Agenda

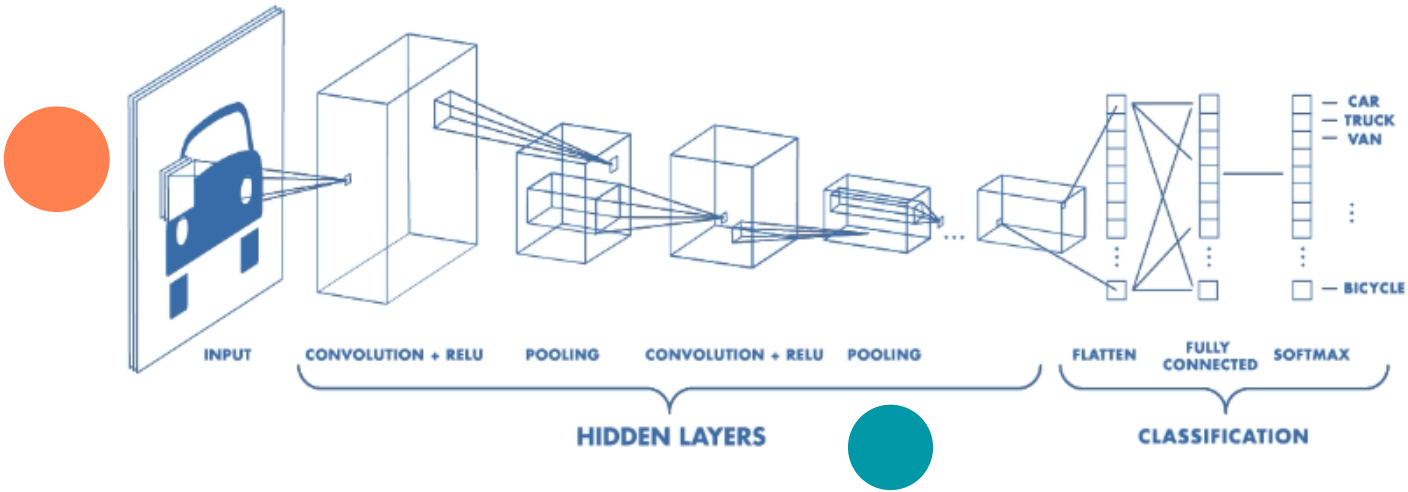
01 Image Classification (CNN)

02 Object Detection (YOLO)

03 Master’s Thesis Research

Image Classification (CNN)

Convolutional Neural Network (CNN)



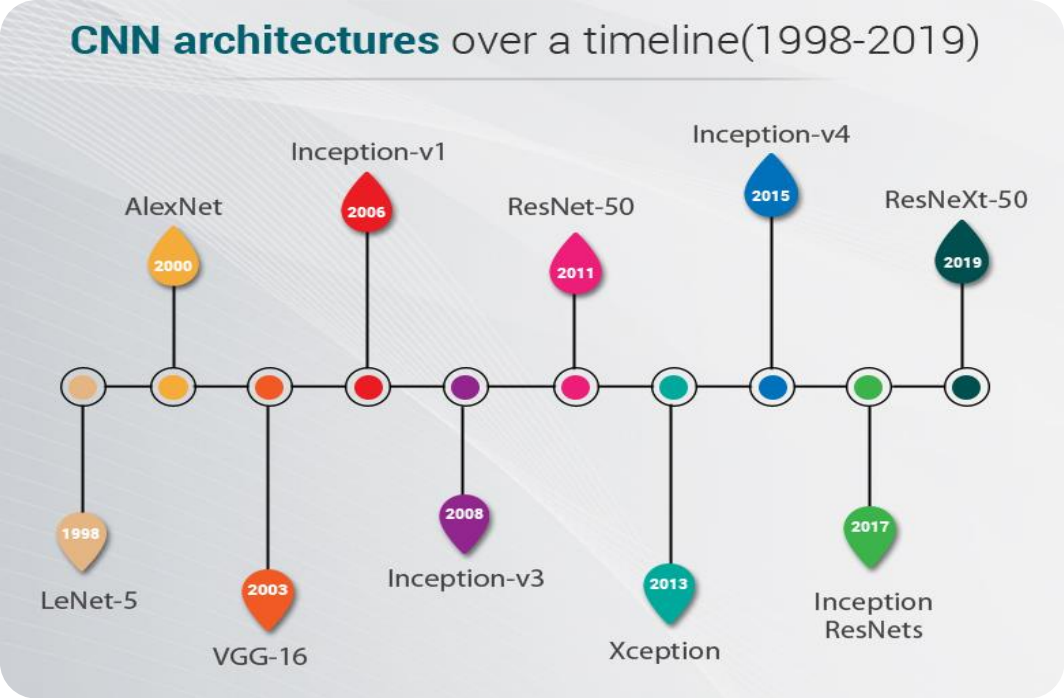
CNN History

Yann Lecun

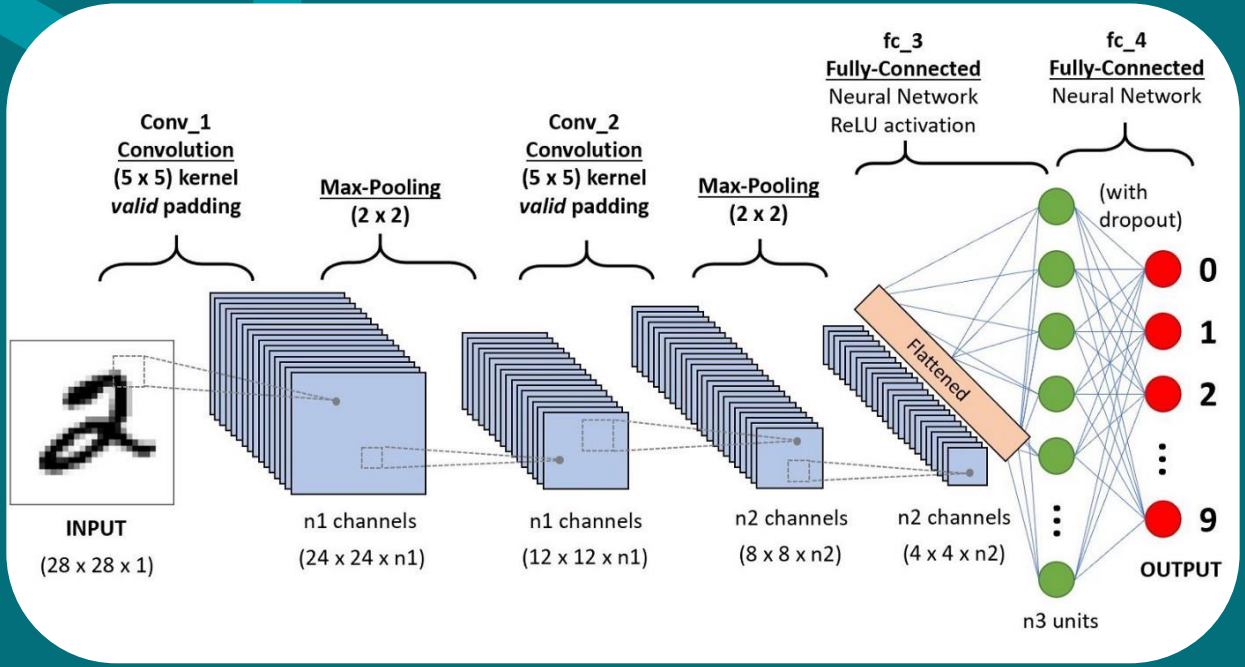
Pionir CNN, Direktor Facebook AI Research Group

Pembuat arsitektur CNN pertama yang disebut **LeNet** pada tahun 1988. Arsitektur tersebut pertama kali digunakan untuk mengenali karakter seperti membaca digit, dan zip codes.

CNN History

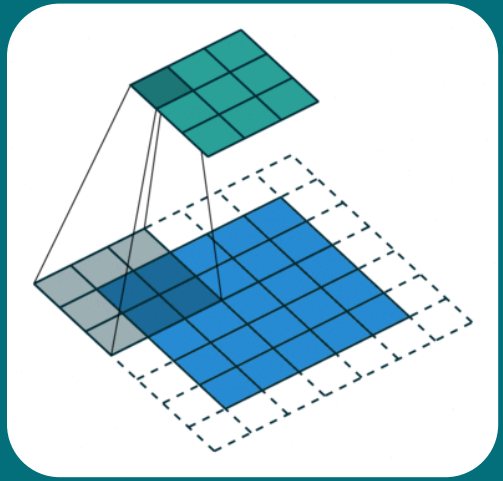

CNN architectures over a timeline(1998-2019)
LeNet-5
1998
AlexNet
2000
VGG-16
2003
Inception-v1
2006
Inception-v3
2008
ResNet-50
2011
Xception
2013
Inception-v4
2015
Inception ResNets
2017
ResNeXt-50
2019

How CNN works?



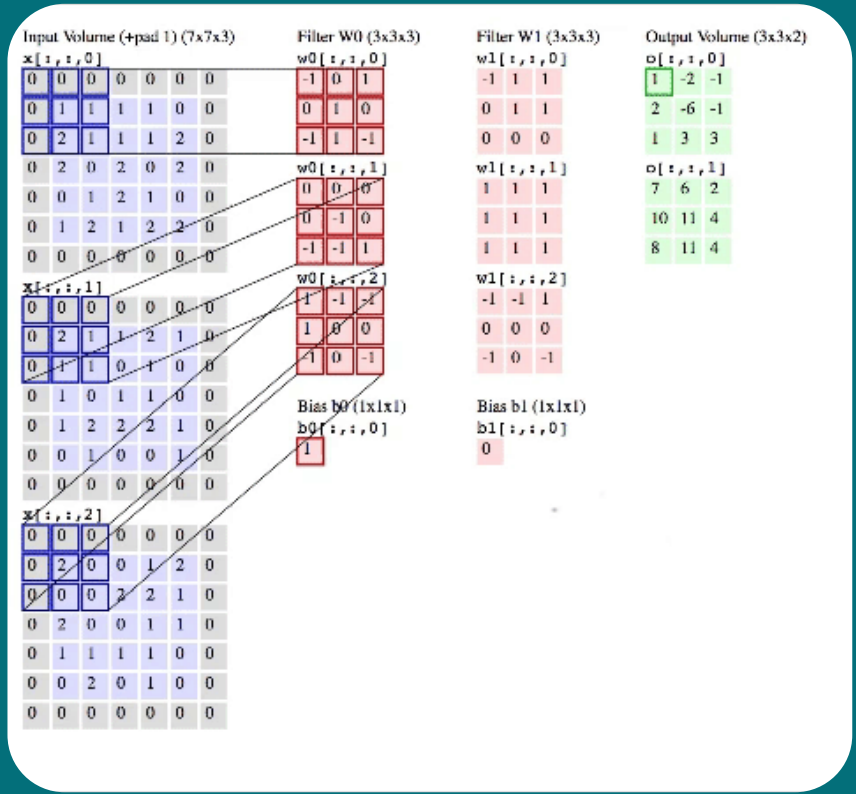
How CNN works?

Convolutional Layer



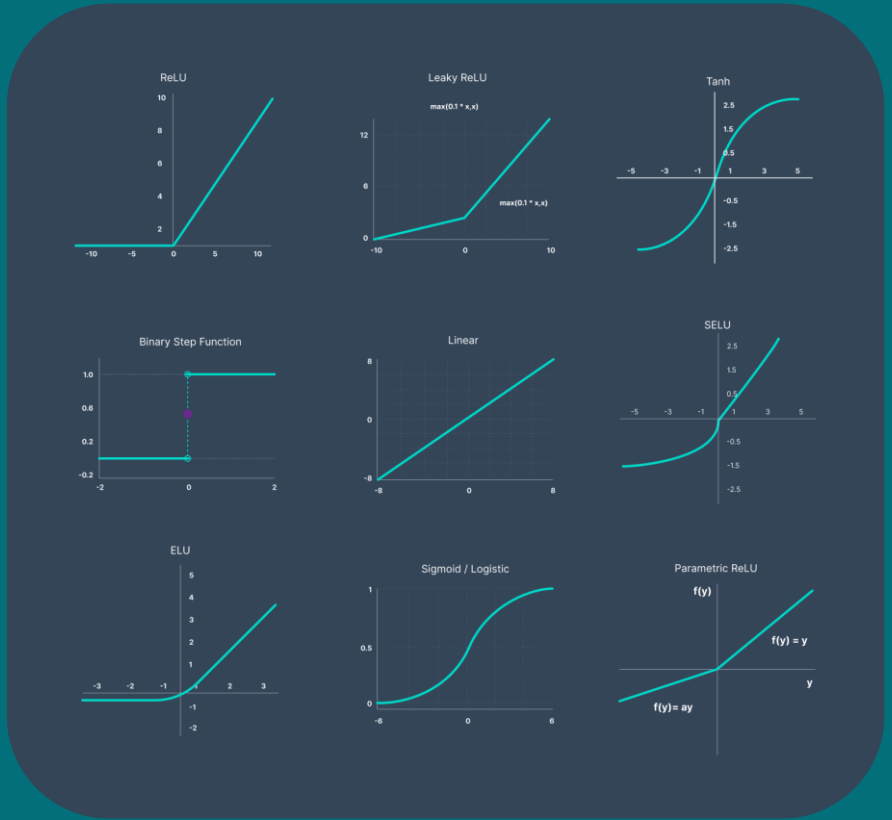
How CNN works?

Convolutional Layer



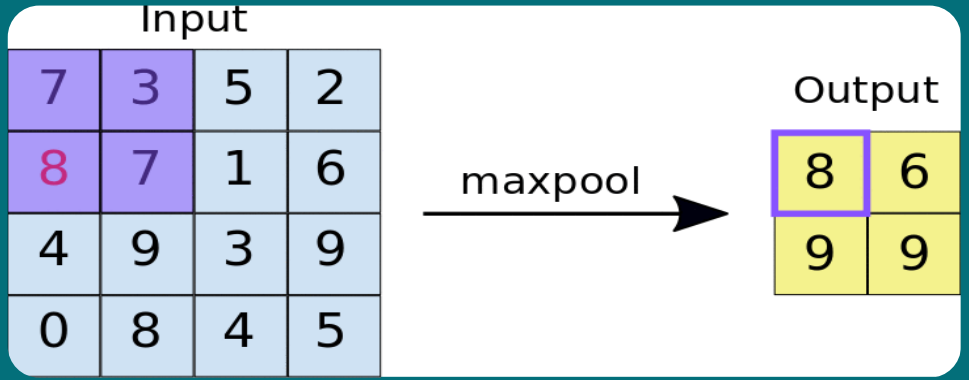
How CNN works?

Activation Function Layer



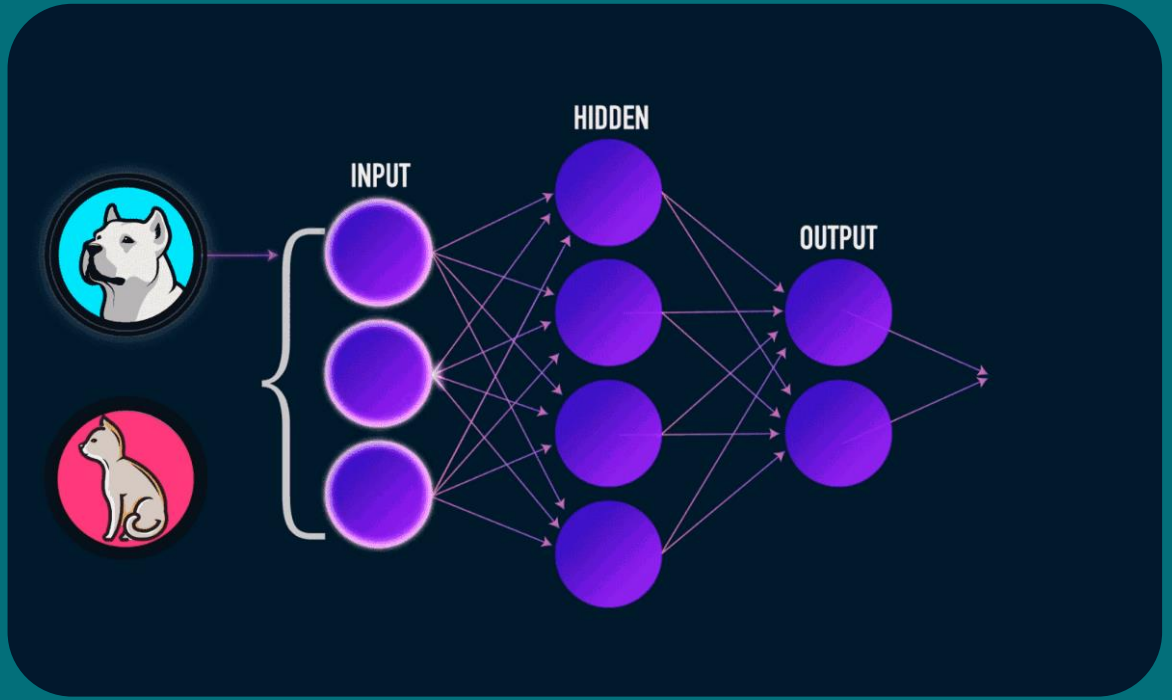
How CNN works?

Pooling Layer



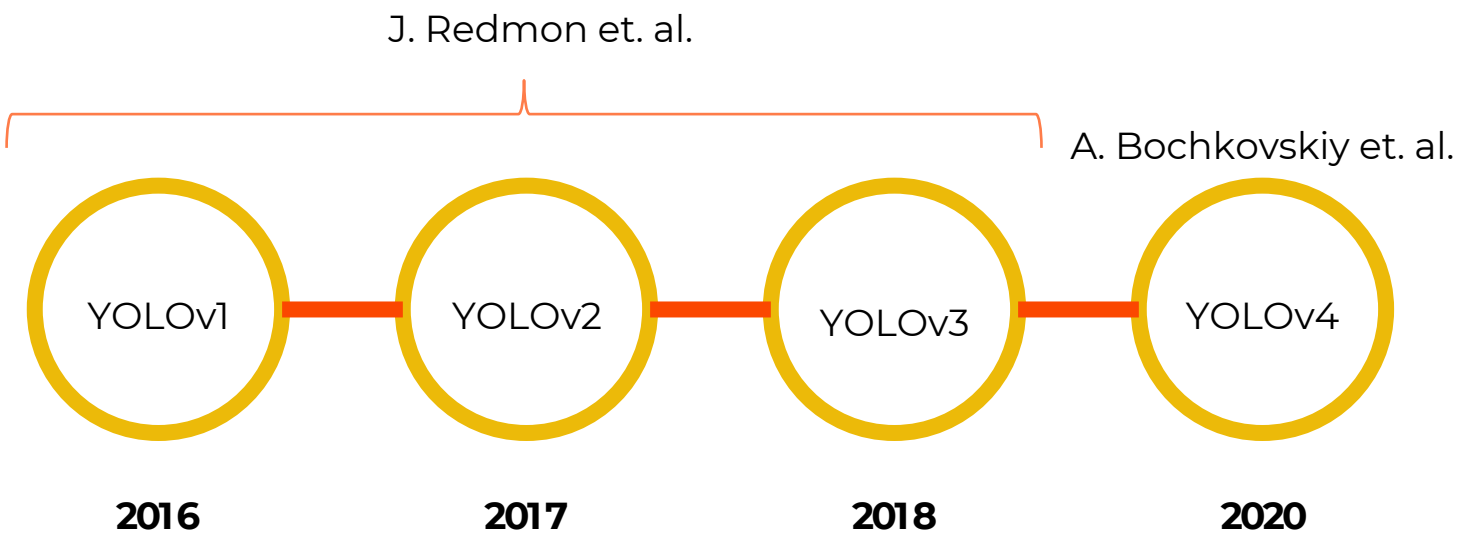
How CNN works?

Artificial Neural Network (ANN)

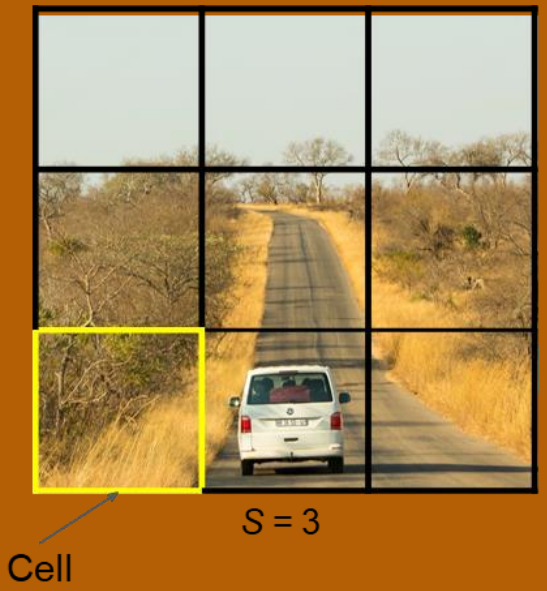


Object Detection (YOLO)

YOLO History

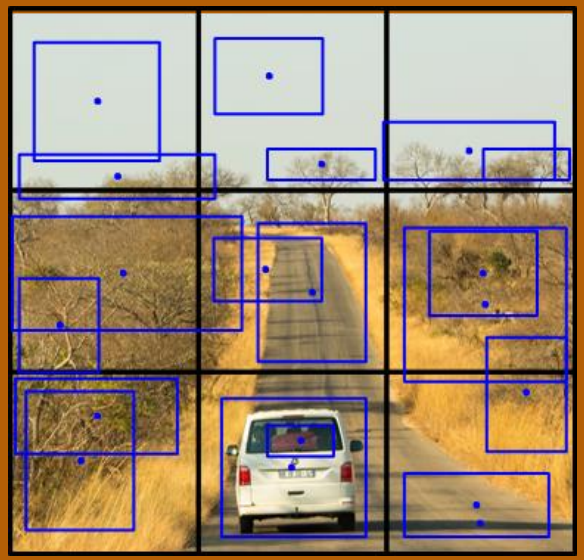


How YOLO works?



1. Gambar dibagi menjadi SxS grid

How YOLO works?



2. Setiap cell dalam grid akan menghasilkan sejumlah B bounding boxes

B = 2

How YOLO works?

3. Predicted bounding boxes akan dipilih sesuai confidence score paling tinggi

YOLO Advantages

Metode	Kecepatan Deteksi
RCNN	0,05 FPS (20 d/g)
Fast RCNN	0,5 FPS (2 d/g)
Fastest RCNN	7 FPS (140 md/g)
YOLO	45 FPS (22 md/g)

d/g = detik per gambar
md/g = milidetik per gambar

YOLO Versions Improvement

YOLOv2

Penghilangan Fully-Connected Layers

YOLOv3

Perbaikan/modifikasi feature extractors.

YOLOv4

Penambahan mekanisme BoF (Bag of Freebies) dan BoS (Bag of Specials)

Masked Face Recognition using Deep Learning based on Unmasked Area (Master's Thesis)

Latar Belakang

Performansi sistem pengenalan wajah bermasker yang belum maksimal

Teknik training penelitian sebelumnya yang kurang tepat

Dataset wajah bermasker yang masih sedikit

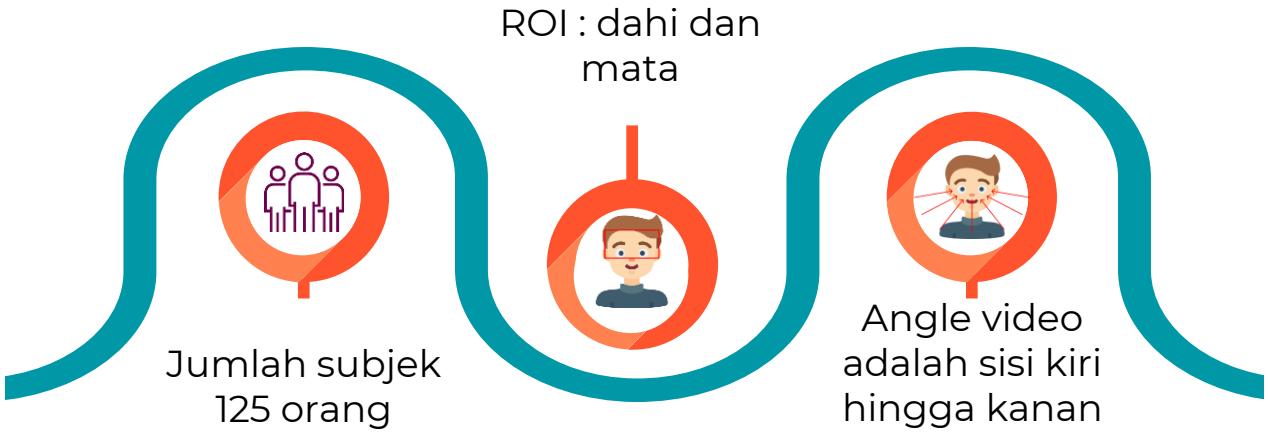
Rumusan Masalah

Bagaimana mendeteksi dan mengenali wajah pada wajah bermasker.

Tujuan Penelitian

- Membangun dataset dengan wajah bermasker dan tidak bermasker berbasis video.
- Mendeteksi dan mengenali wajah pada wajah bermasker.
- Mengevaluasi kinerja sistem pengenalan wajah bermasker.

Batasan Masalah



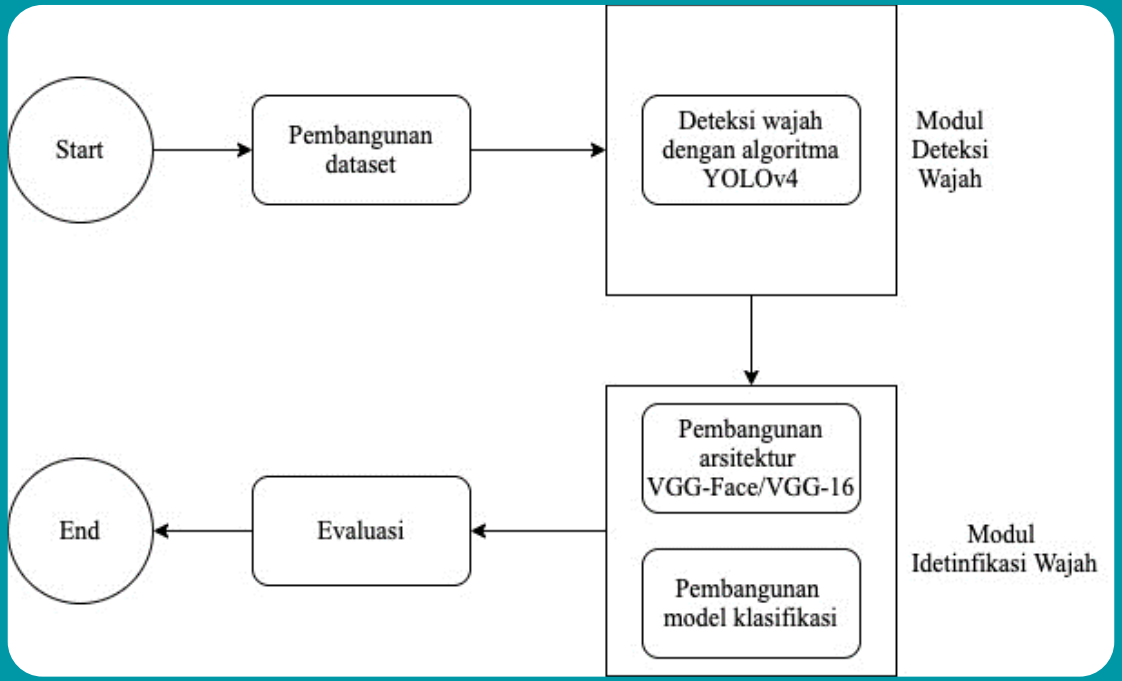
Penelitian Rujukkan

Elmahmudi et. al. (2018)	Experiments on Deep Face Recognition Using Partial Faces
Elmahmudi et. al. (2019)	Deep face recognition using imperfect facial data
Mundial et. al. (2020)	Towards Facial Recognition Problem in COVID-19 Pandemic

Mundial et. al. **(2020)**

Jenis Pengujian	Dataset	Akurasi
Testing	LFW Dataset (training dengan wajah bermasker)	98%
Testing	Dataset sendiri (training tanpa wajah bermasker)	79%
Testing	Dataset sendiri (training dengan wajah bermasker)	97%
Testing	RWMFD (dengan training wajah bermasker)	97%

Alur Penelitian



Pembangunan Dataset

Jumlah subjek terkumpul : 125

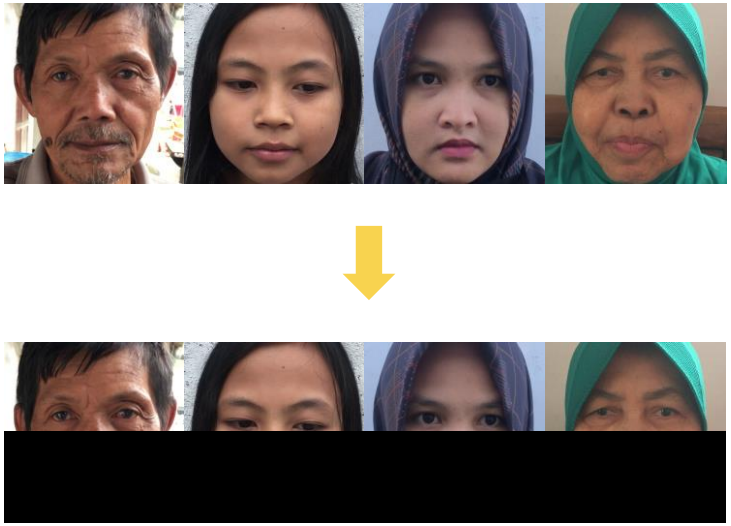
Range umur : anak-anak sampai lansia

Durasi : 5 – 7 detik

Jumlah video per subjek : 2

Tools/device : Kamera smartphone

Modul Deteksi Wajah & Preprocessing



Training

Testing

Modul Identifikasi Wajah

1. 80% data tidak bermasker untuk data train, 20% lainnya untuk validasi.

2. Total frame yang dimuat setiap video : 20%, 30% dan 40%

Train Data Size : 3005	Train Data Size : 4527	Train Data Size : 6059
Validation Data Size : 752	Validation Data Size : 1132	Validation Data Size : 1515
Test Data Size : 3829	Test Data Size : 5762	Test Data Size : 7710
20%	30%	40%

3. Eksperimen:
   - Baseline
   - Teknik training penelitian sebelumnya (Mundial dkk. 2020)
   - VGG-16 (non pre-train)
   - VGG-Face (pre-train)

VGG-16

Layer	Jumlah Layer	Ukuran Tiap Layer
Input	1	224x224x3
Conv1	2	224x224x64
Pool 1	1	112x112x64
Conv2	2	112x112x128
Pool 2	1	56x56x128
Conv3	3	56x56x256
Pool 3	1	28x28x256
Conv4	3	28x28x512
Pool 4	1	14x14x512
Conv5	3	14x14x512
Pool 5	1	7x7x512
FC 6	1	1x1x4096
FC 7	1	1x1x4096
FC 8	1	1x1x1000
Softmax	1	1x1x125

VGG-Face

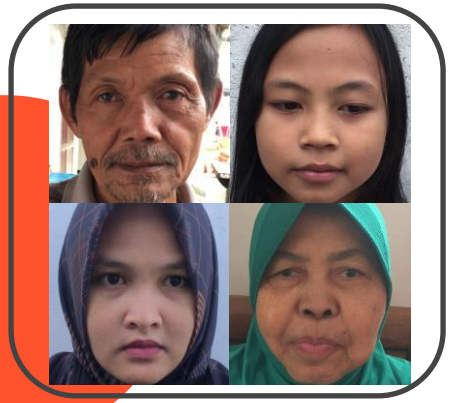
Layer	Jumlah Layer	Ukuran Tiap Layer
Input	1	224x224x3
Conv1	2	224x224x64
Pool 1	1	112x112x64
Conv2	2	112x112x128
Pool 2	1	56x56x128
Conv3	3	56x56x256
Pool 3	1	28x28x256
Conv4	3	28x28x512
Pool 4	1	14x14x512
Conv5	3	14x14x512
Pool 5	1	7x7x512
FC 6	1	1x1x4096
FC 7	1	1x1x4096
FC 8	1	1x1x2622

Baseline Program

- Dataset yang digunakan: FEI-Dataset (200 individu, 2800 citra)
- Pendekatan : VGG-Face
- Partisi Data : 20% data test, 80% data train
- Classifier : Cosine Similarity (CS)

Skenario	Classifier	Akurasi Training	Akurasi Testing
Full 1 wajah	CS	100%	100%
Setengah bagian atas wajah	CS	100%	98.56%

Mundial et. al. (2020)



Training

Testing

Mundial et. al. (2020)

ANN

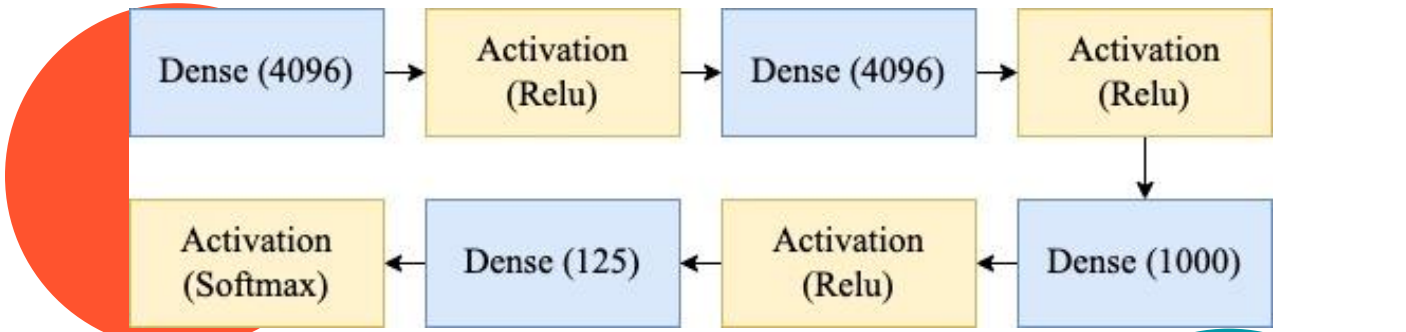
Total Frame	Akurasi Training	Akurasi Validasi	Akurasi Testing	Precision	Recall	F1-Score
20%	100%	100%	79.58%	91.30%	79.58%	76.97%
30%	100%	100%	77.02%	88.40%	77.02%	73.66%
40%	100%	100%	75.03%	87.87%	75.03%	70.51%

CS

Total Frame	Akurasi Training	Akurasi Validasi	Akurasi Testing	Precision	Recall	F1-Score
20%	100%	100%	79.58%	81.03%	79.58%	76.20%
30%	100%	100%	78.98%	78.83%	78.98%	75.32%
40%	100%	100%	79.26%	80.35%	79.26%	75.58%

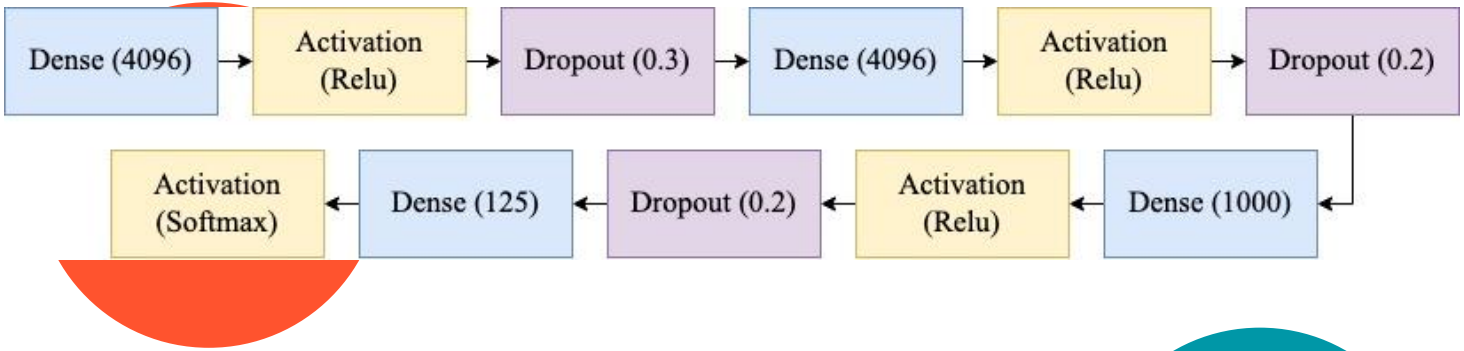
Eskperimen VGG-16

Konfigurasi 1



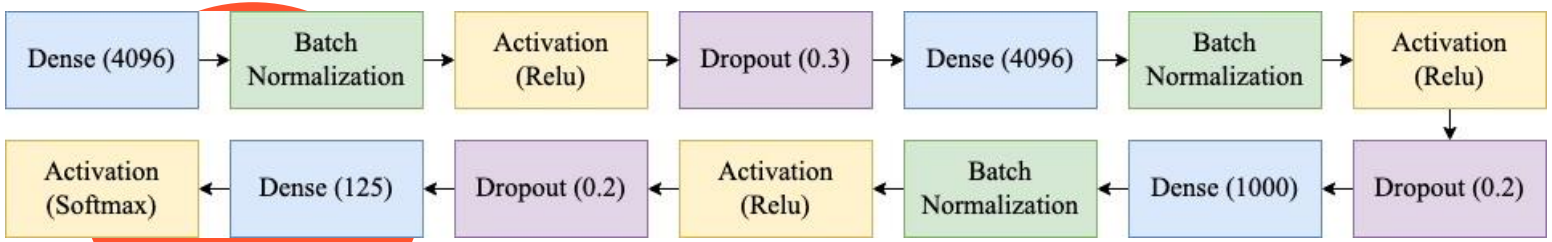
Eskperimen VGG-16

Konfigurasi 2

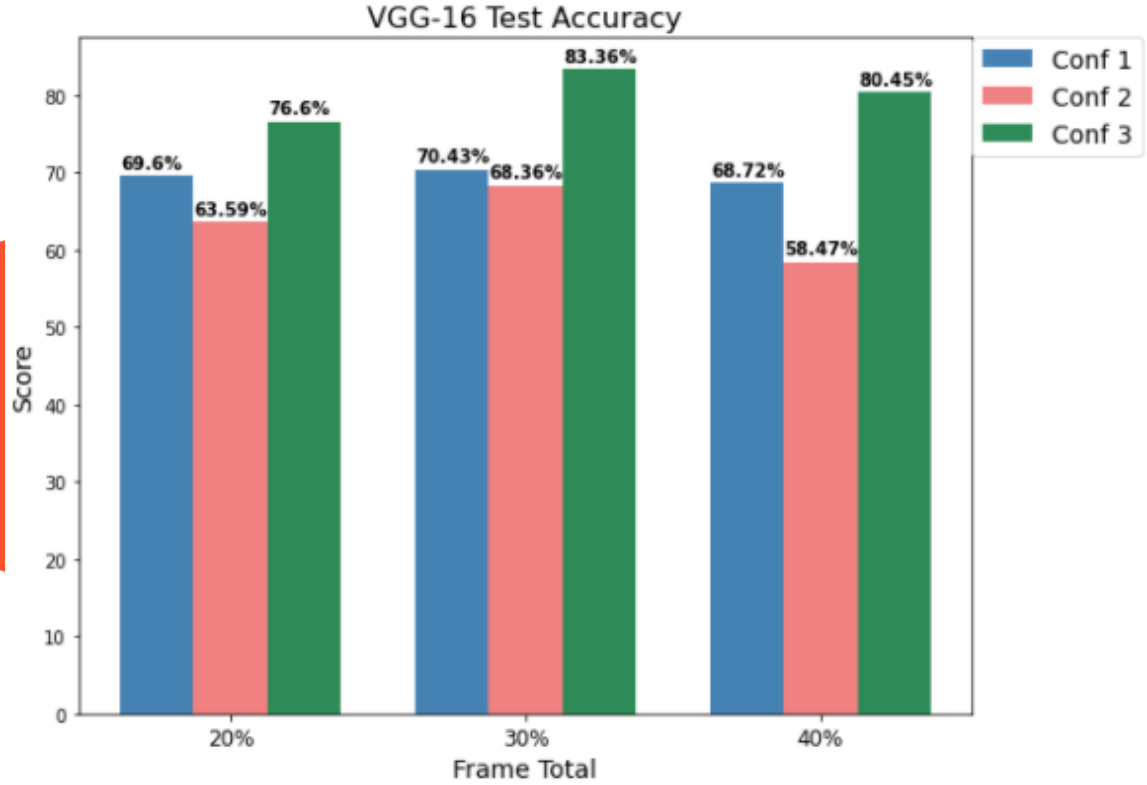


Eskperimen VGG-16

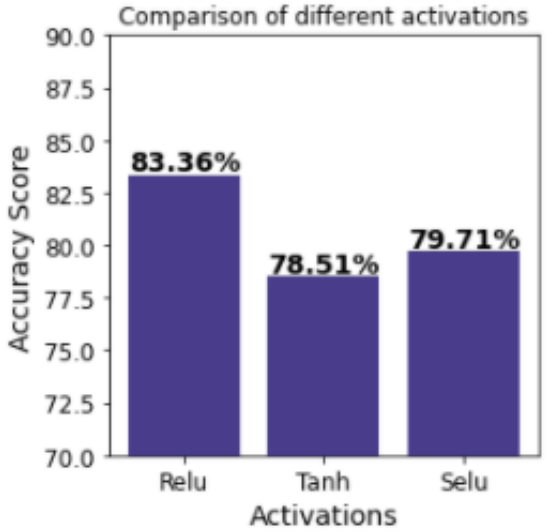
Konfigurasi 3



Eskperimen VGG-16

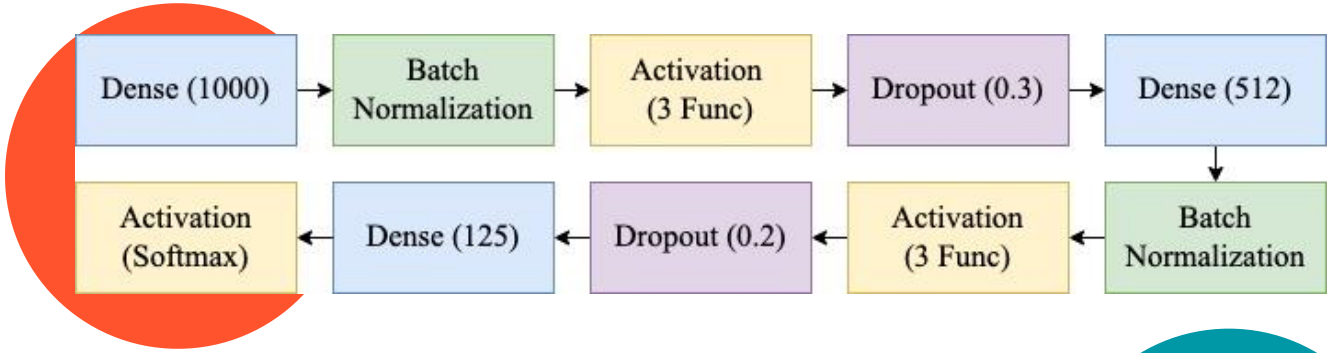


Eskperimen VGG-16

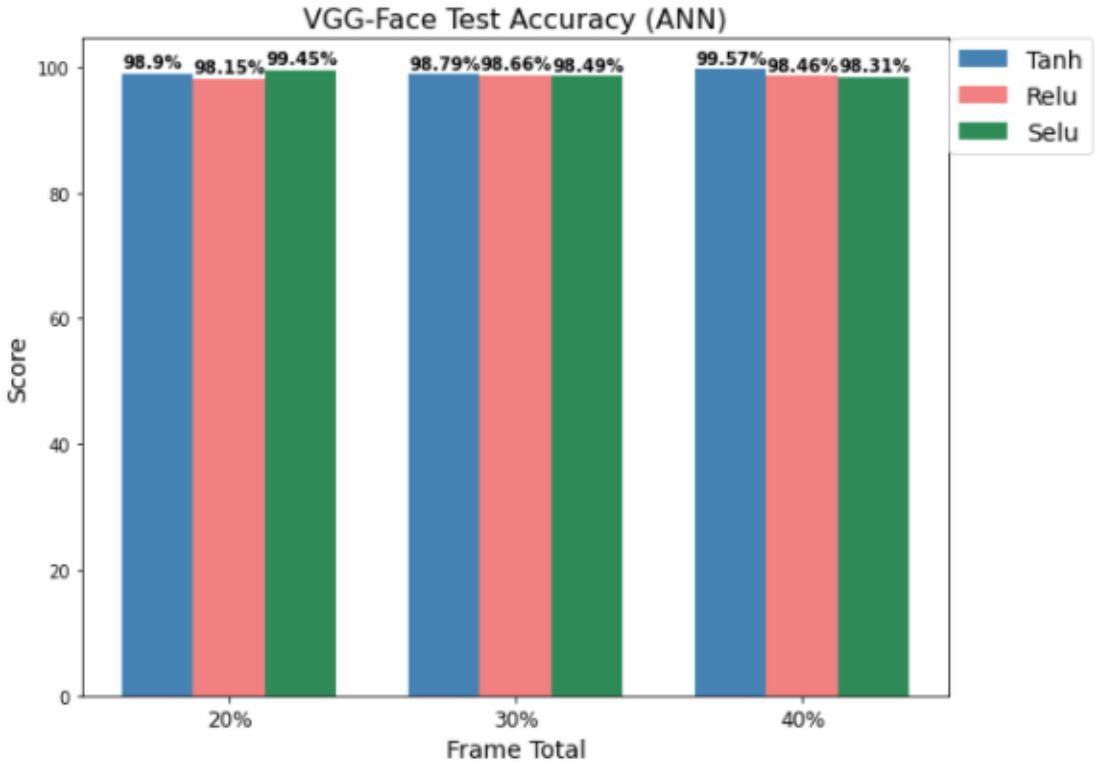


Eskperimen VGG-Face

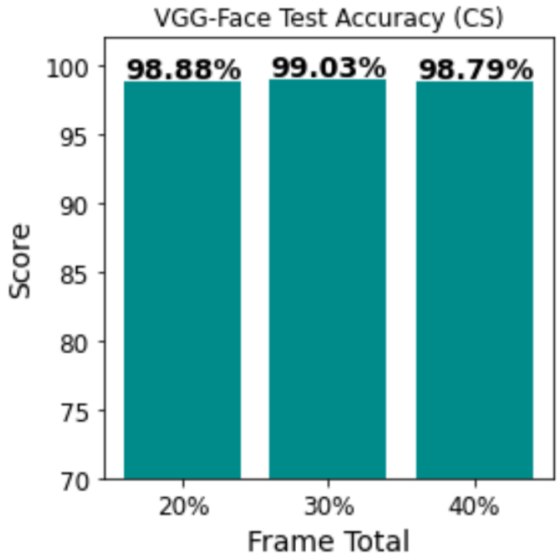
Konfigurasi ANN



Eskperimen VGG-Face



Eskperimen VGG-Face



Rekap Performansi

Teknik	Total Frame	Akurasi Training	Akurasi Validasi	Akurasi Testing
Mundial dkk. (2020)	20%	100%	100%	79.58%
VGG-16 (Conf 1)	30%	100%	100%	70.43%
VGG-16 (Conf 2)	30%	100%	100%	68.36%
VGG-16 (Conf 3, relu)	30%	100%	100%	83.36%
VGG-16 (Conf 3, tanh)	30%	100%	100%	78.51%
VGG-16 (Conf 3, selu)	30%	100%	100%	79.71%
VGG-Face (ANN)	40%	100%	100%	99.57%
VGG-Face (CS)	30%	100%	100%	99.03%

Kesimpulan & Saran

Kesimpulan

- Dataset telah dibangun dengan jumlah subjek sebanyak 125 orang, dengan format video.
- Model terbaik yang dibangun adalah VGG-Face dengan classifier ANN serta fungsi aktivasi Tanh. Akurasi uji yang didapatkan sebesar 99,57% (Dengan total frame 40%)
- Eksperimen real-time untuk konsep surveillance system belum memuaskan.

Saran

- Mencari model non-pre train yang tidak overfit.
- Menyempurnakan model identifikasi.
- Penambahan subjek dataset.

DEMO

Thanks!