

Laporan Akhir Penelitian Smart CCTV Makassar

**Optimalisasi Algoritma Computer Vision Terhadap Kondisi
Lingkungan Kota Makassar Untuk Mendeteksi Orang Tidak
Dikenal Berbasis Video CCTV**



PENELITIAN, PENGEMBANGAN DAN PEREKAYASAAN DI BIDANG
INOVASI DAN TEKNOLOGI - Badan Penelitian dan Pengembangan Daerah
Kota Makassar

Disusun Oleh:

Prof. Dr. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN.ENG

Nublan Azqalani S.T

Muhammad Fadhil Bahrin S.T

DEPARTEMEN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023

DAFTAR ISI

DAFTAR ISI	2
DAFTAR GAMBAR.....	4
DAFTAR TABEL	6
BAB I PENDAHULUAN.....	7
1.1. Latar Belakang	7
1.2. Rumusan Masalah:.....	11
1.3. Tujuan Penelitian	11
1.3. Manfaat Penelitian	12
1.4. Batasan Masalah	12
1.5. Sistematika Penulisan.....	13
BAB II TINJAUAN PUSTAKA	14
2.1. Landasan Teori	14
2.2. Penelitian Terkait	57
2.3. State Of The Art	60
BAB III.....	62
3.1. Tahapan Penelitian.....	62
3.2. Waktu dan Lokasi Penelitian.....	63
3.3. Populasi dan Sampel	63
3.4. Teknik Pengambilan data	64
3.5. Desain Usulan sistem	65
3.6. Perangkat Pengiriman Video	69
3.7. Teknik Analisa Kinerja Sistem	72
3.8. Instrumen Penelitian.....	73
BAB IV HASIL DAN PEMBAHASAN	75
4.1. Tahap Pengambilan Data.....	75
4.2. Pre-processing dan Augmentasi Data.....	76

4.3.	Tahap Pembangunan Model GhostFaceNet.....	78
4.4.	Tahap Pengujian Model GhostFaceNet.....	83
4.5.	Tahap Pengujian Berdasarkan Data Video	86
4.6.	Tahap Integrasi GhostFaceNet – DeepSORT	94
4.7.	Pembuatan ROI Wilayah Sekolah	99
4.8.	Database Management	101
4.9.	Menghubungkan Kamera IP dan RaspberryPi	104
4.10.	Pengiriman Video ke Server melalui Koneksi Socket	105
4.11.	Notifikasi WhatsApp.....	108
4.12.	Analisis Kebutuhan Infrastruktur.....	109
BAB V KESIMPULAN DAN SARAN		112
5.1	Kesimpulan	112
5.2	Saran	113

DAFTAR GAMBAR

Gambar 2. 1 Rangkaian umum proses computer vision	18
Gambar 2. 2 Ilustrasi Hasil deteksi wajah.....	32
Gambar 2. 3 Contoh hasil rekognisi wajah	33
Gambar 2. 4 Beberapa arsitektur GhostFaceNets bekerja	51
Gambar 2. 5 Alur kerja DeepSORT	57
Gambar 3. 1 Alur Penelitian	62
Gambar 3. 2 Gambaran Teknik Pengambilan data	64
Gambar 3. 3 Contoh ilustrasi pengaturan kamera.....	64
Gambar 3. 4 Desain Umum Sistem.....	65
Gambar 3. 5 Proses Pembuatan Model	68
Gambar 4. 1 Teknik Pengambilan Data	75
Gambar 4. 2 Struktur Penyimpanan data	75
Gambar 4. 3 Contoh Data Asli Hasil Pengambilan Data	76
Gambar 4. 4 Keseluruhan proses pre-processing dan augmentasi pada data.....	77
Gambar 4. 5 Data hasil pre-processing dan augmentasi	Error! Bookmark not defined.
Gambar 4. 6 Data Format .bin.....	84
Gambar 4. 7 Contoh Grafik akurasi hasil pembelajaran pada tiap iterasi	85
Gambar 4. 8 Visualisasi 7 layer awal pada mode yang terbentuk	87
Gambar 4. 9 Beberapa contoh hasil pendeteksian oleh sistem pada scenario 1	91
Gambar 4. 10 Beberapa contoh hasil pendeteksian oleh sistem pada scenario 2	92
Gambar 4. 11 Hasil pendeteksian oleh sistem pada scenario 3.....	93
Gambar 4. 12 Beberapa contoh hasil deteksi + recognisi + tracking oleh sistem pada scenario 1	97
Gambar 4. 13 Beberapa contoh hasil deteksi + recognisi + tracking oleh sistem pada scenario 2	98

Gambar 4. 14 Beberapa contoh hasil deteksi + recognisi + tracking oleh sistem pada scenario 3	99
Gambar 4. 15 Hasil pembuatan ROI pada sistem	101
Gambar 4. 16 Skema jaringan pada sistem	105

DAFTAR TABEL

Tabel 2. 1 State of The Art	60
Tabel 4. 1 Tabel Analisis Sistem Menggunakan Video CCTV Pada Skenario 1	89
Tabel 4. 2 Tabel Analisis Sistem Menggunakan Video CCTV Pada Skenario 2	91

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sebuah kota yang berkualitas harus dapat menyediakan kenyamanan dan keamanan yang optimal bagi penduduk yang tinggal di dalamnya. Konsep sebuah kota yang nyaman dan aman sebagai lingkungan tempat tinggal dan beraktivitas membutuhkan adanya inovasi teknologi. Hal ini dikarenakan lingkungan kota yang nyaman dan aman harus dapat memenuhi kebutuhan dari berbagai aspek, baik itu aspek fisik infrastruktur maupun aspek sosial. Dalam menjalankan tanggung jawabnya sebagai pemerintah eksekutif di kota Makassar, walikota dan wakil walikota Makassar periode 2019-2024 telah dengan tegas menggarisbawahi visi dan misi mereka untuk mewujudkan konsep kota yang aman dan nyaman, terutama dalam misi ketiga yang diusung, yaitu 'Restorasi Ruang Kota yang inklusif menuju kota nyaman kelas dunia "Sombere & Smart" City untuk Semua' (*Visi Misi - BAPPEDA Makassar, n.d.*).

Diantara kelompok umur di elemen masyarakat kota, kategori anak-anak seringkali kurang mendapatkan perhatian lebih terutama dalam upaya menciptakan lingkungan yang aman dan nyaman. Berdasarkan data yang dirilis Dinas Perlindungan Perempuan dan Anak (DP3A) Makassar, terdapat 488 kasus kekerasan anak dan perempuan di Makassar pada tahun 2022. Dari angka itu 200 kasus merupakan kekerasan terhadap anak. Kekerasan terhadap anak menempati urutan pertama dalam pencatatan DP3A Makassar (Rin, n.d.). Di posisi kedua anak berhadapan hukum 100 kasus, selanjutnya kekerasan terhadap perempuan, KDRT

dan kasus lainnya. Berbagai upaya telah dilakukan untuk menekan angka tersebut, melalui program “Jagai Anakta’ “(diskominfo, 2023), pemerintah kota makassar mengajak untuk berkolaborasi dengan sebanyak mungkin elemen masyarakat untuk menuntaskan masalah ini. Salah satu wilayah penting yang perlu mendapatkan perhatian adalah wilayah persekolahan, durasi waktu antara anak pulang sekolah hingga selamat sampai dirumah dinilai masih menjadi fase ‘critical time’ terjadinya kejahatan terhadap anak dikarekanak pengawasan orang dewasa mulai mengalami kelonggaran. Oleh karena itu, diperlukan penggunaan CCTV yang dilengkapi dengan kecerdasan buatan (artificial intelligence) untuk membantu menjaga keamanan dan ketertiban di area gerbang sekolah.

Pemasangan kamera CCTV dibeberapa titik di kota makassar dalam upaya pengawasan telah berjalan sejak dicanangkannya program MAKAVERSE oleh walikota dan wakil walikota makassar. Namun, cctv yang ada masih belum dibekali dengan kemampuan artificial intelligence yang dapat secara otomatis mendeteksi perilaku masyarakat di wilayah yang dimaksudkan, hal ini menyebabkan pengawasan masih bersifat semi-manual yang kurang efektif dan menyeluruh.

Penerapan kamera cctv yang dibekali AI sendiri bukan merupakan bidang penelitian baru. Dalam penelitiannya yang berjudul ‘Experimental Analysis of Face Recognition on Still and CCTV images’, Shaokang Chen dkk menyimpulkan bahwa pengenalan wajah berbasis fitur holistik yang dapat mentolerir adanya degradasi citra namun sangat sensitif terhadap transformasi geometris (Chen et al., 2008). Sehingga dapat dijadikan catatan, bahwa untuk penggunaan artificial intelligence pada CCTV memerlukan pengambilan data dari banyak sudut pandang agar tindakan transformasi

geometris pada sekumpulan data latih dapat di tiadakan dan fitur holistic dari objek yang ingin di deteksi menjadi lebih kaya. Hal ini menyebabkan focus peneltian dengan kasus yang mirip dapat berfokus pada proses-proses augmentasi yang menyesuaikan dengan kemungkinan-kemungkinan keadaan lingkungan yang dapat mengganggu proses ekstraksi fitur. Perubahan keadaan lingkungan seperti perubahan intensitas cahaya dan adanya hujan menjadi lebih relevan untuk diteliti lebih jauh dengan berusaha mengoptimalkan algoritma computer vision yang sudah ada.

Selain masalah keadaan lingkungan yang dapat mengganggu proses deteksi dan recognisi, adanya oklusi yang dapat menghalangi fitur-fitur di wajah juga seringkali menjadi focus dalam penelitian mengenai penggunaan cctv untuk alasan pemantauan keamanan. Dalam penelitiannya mengenai 'Face Occlusion Recognition With Deep Learning in Security Framework for the IoT' pada tahun 2019, Li Mao dkk mengusulkan untuk menggabungkan metode Gaussian energy-based untuk deteksi dan lokalisasi kepala pada manusia, CNN sebagai algoritma untuk ekstraksi fitur wajah dan metode Sparse Model-Based Classifier untuk melakukan klasifikasi berdasarkan fitur yang telah dihitung oleh CNN. Metode ini terbukti mendapatkan akurasi 98,89% pada kasus deteksi wajah dengan atau tanpa oklusi wajah seperti kacamata dan masker dengan menggunakan perangkat rendah komputasi(Mao et al., 2019).

Dalam lingkup Universitas Hasanuddin sendir, Muhammad Arafah telah melakukan penelitian terkait penentuan jarak terbaik dalam proses instalasi CCTV dalam mengidentifikasi wajah seseorang di area pemeriksaan penumpang di bandara. Menggunakan serangkaian pre-processing data seperti Brightness Enhancement

(BE), Grayscale dan Contrast Limited Adaptive Histogram Equalization (CLAHE) untuk meningkatkan kualitas wajah yang masuk kedalam frame, kemudian menggunakan metode Viola-Jones untuk proses ekstraksi fitur wajah dan SVM untuk klasifikasi wajah mendapatkan hasil bahwa jarak paling optimal adalah jarak individu terhadap CCTV adalah 300 cm dengan posisi CCTV berada 250 cm dari permukaan tanah dengan tingkat akurasi klasifikasi wajah sebesar 86,76% (Arafah et al., 2019).

Untuk menyelesaikan masalah pendeteksian orang tidak dikenal di wilayah sekolah di kota Makassar menggunakan algoritma-algoritma computer vision terdapat tiga proses utama yaitu; bagaimana mendapatkan algoritma deteksi wajah dengan beban komputasi seminimal mungkin dengan akurasi yang paling optimal, bagaimana mendapatkan algoritma tracking yang nantinya dapat mendeteksi keberadaan setiap individu yang telah dinyatakan telah terdeteksi (apakah masih berada di wilayah sekolah atau tidak) dan masalah terakhir adalah bagaimana mengirimkan notifikasi secara real-time mengenai posisi siswa.

Dalam kasus deteksi wajah sudah terdapat beberapa algoritma dan metode dengan variasi benchmarking waktu komputasi dan akurasi seperti algoritma FaceNet, ArcFace, DisFace, GhostFaceNet, dan beberapa algoritma deteksi dan pengenalan wajah lainnya. Untuk algoritma tracking sendiri sudah dikenal Deep SORT, Kalman Filter, Particle Filter, Multiple Hypothesis Tracking, dan beberapa algoritma tracking dan pelacakan objek lainnya.

Berdasarkan uraian-uraian diatas untuk mengembangkan system deteksi orang tidak dikenal di area sekolah khususnya dengan kondisi kota Makassar yang

dapat bekerja dengan waktu komputasi yang cepat dengan tingkat akurasi yang tinggi, penelitian ini akan berfokus pada pembuatan model deteksi, recognisi dan tracking wajah individu dalam wilayah sekolah yang terintegrasi dengan system peringatan yang secara real-time akan diterimad di gawai pihak orang dewasa yang bertanggung jawab diwilayah sekolah.

1.2. Rumusan Masalah

Bagaimana mengembangkan system deteksi dan recognisi multi-wajah di wilayah sekolah di kota makassardengan camera cctv secara real-time?

1. Bagaimana mengembangkan system tracking multi-wajah dengan camera cctv diwilayah sekolah di kota makassar secara real-time?
2. Bagaimana membuat system peringatan dini keamanan wilayah sekolah berbasis informasi cctv?
3. Bagaimana unjuk kinerja system deteksi, recognis, tracking wajah di wilayah sekolah di kota makassar?

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini ialah sebagai berikut:

1. Mengembangkan sistem deteksi dan pengenalan multi-wajah secara real-time di wilayah sekolah di Kota Makassar dengan menggunakan kamera CCTV.
2. Mengembangkan sistem pelacakan multi-wajah secara real-time dengan menggunakan kamera CCTV di wilayah sekolah di Kota Makassar.
3. Membuat sistem peringatan dini keamanan wilayah sekolah berbasis informasi dari rekaman CCTV.
4. Mengetahui kinerja sistem deteksi, pengenalan, dan pelacakan wajah di wilayah sekolah di Kota Makassar.

1.3. Manfaat Penelitian

Adapun manfaat dari penelitian ini ialah sebagai berikut:

1. Memberikan informasi terhadap bidang ilmu pengetahuan yang relevan sesuai tema penelitian
2. Membantu menyediakan purwarupa sistem deteksi Orang Tidak Dikenal di Area Sekolah realtime
3. Membantu pemerintah kota Makasaar mewujudkan program smart city
4. Membantu pihak sekolah untuk melakukan pengawasan terutama di fase 'critical time' saat pulang sekolah.

1.4. Batasan Masalah

1. Data latih yang digunakan merupakan data primer yang diambil dilokasi penelitian berupa foto wajah dari beberapa sudut pengambilan
2. System akan diuji menggunakan data streaming yang berasal dari CCTV yang terpasang di sekolah.
3. Proses augmentasi data akan menyesuaikan dengan kondisi sekolah, seperti range intensitas cahaya, penambahan filter jika terjadi hujan
4. Oklusi yang termasuk pada penelitian ini adalah penggunaan kerudung.
5. Sistem yang dibangun belum dilatih untuk mengatasi oklusi masker dan kacamata hitam, hal ini dapat dilakukan dengan melakukan pengambilan data ulang
6. Luas ruang yang direkam pada penelitian ini adalah 4x7 meter

1.5. Sistematika Penulisan

Berikut adalah gambaran singkat mengenai isi tulisan secara keseluruhan:

BAB 1 PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah penelitian, metode penelitian, serta sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada bab ini membahas landasan teori yang digunakan untuk menganalisis masalah yang akan diteliti serta hal-hal lain yang berhubungan dengan variabel-variabel data yang akan digunakan, Computer Vision, Face Detection, Face Recognition, maupun Surveillance Detection System.

BAB 3 METODOLOGI PENELITIAN

Pada bab ini berisi mengenai tahap penelitian, instrumen penelitian, pengumpulan data, penerapan algoritma, teknik pengolahan data.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini berisi tentang sistem yang telah berhasil dibangun serta pembahasannya.

BAB 5 PENUTUP

Bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut

BAB II

TINJAUAN PUSTAKA

2.1. Landasan Teori

2.1.1. Artificial Intelligence

Pada tahun 1956, dalam sebuah konferensi di Universitas Dartmouth, para ilmuwan secara resmi mengusulkan istilah "kecerdasan buatan" (artificial intelligence). Momen ini merupakan langkah pertama dalam penelitian topik baru tentang bagaimana mesin bisa mensimulasikan aktivitas intelektual manusia. Pada awal tahun 2016, AlphaGo berhasil mengalahkan juara catur dunia. Kejadian ini segera membangkitkan minat global terhadap kecerdasan buatan (AI). Pengembangan kecerdasan buatan telah memberikan manfaat ekonomi besar bagi manusia dan telah bermanfaat dalam semua aspek kehidupan, serta telah secara besar-besaran mendorong perkembangan sosial dan membawa masyarakat ke era baru. Banyak ilmuwan mulai melakukan penelitian terkait AI sejak akhir abad ke-20.

AI adalah istilah umum untuk ilmu kecerdasan buatan. Ini menggunakan komputer untuk mensimulasikan perilaku intelektual manusia dan melatih komputer untuk mempelajari perilaku manusia seperti belajar, penilaian, dan pengambilan keputusan. AI adalah proyek pengetahuan yang mengambil pengetahuan sebagai objek, mengakuisisi pengetahuan, menganalisis dan mempelajari metode ekspresi pengetahuan, dan menggunakan pendekatan ini untuk mencapai efek mensimulasikan aktivitas intelektual manusia. AI adalah kompilasi ilmu komputer, logika, biologi, psikologi, filsafat, dan banyak disiplin lainnya, dan telah mencapai hasil yang luar biasa dalam aplikasi seperti

pengenalan suara, pemrosesan gambar, pemrosesan bahasa alami, pembuktian teorema otomatis, dan robot pintar. AI memainkan peran yang tak tergantikan dalam perkembangan sosial dan telah membawa hasil revolusioner dalam meningkatkan efisiensi kerja, mengurangi biaya tenaga kerja, mengoptimalkan struktur sumber daya manusia, dan menciptakan permintaan pekerjaan baru.

Kecerdasan buatan adalah studi tentang bagaimana membuat komputer melakukan tugas-tugas cerdas yang sebelumnya hanya bisa dilakukan oleh manusia. Dalam beberapa tahun terakhir, AI telah berkembang dengan pesat dan telah mengubah gaya hidup manusia. Pengembangan AI telah menjadi strategi pengembangan penting bagi negara-negara di seluruh dunia, meningkatkan daya saing nasional dan menjaga keamanan. Banyak negara telah mengenalkan kebijakan preferensial dan telah memperkuat pengembangan teknologi kunci dan talenta-talenta untuk memimpin dalam kompetisi internasional yang baru. AI telah menjadi pusat penelitian dalam ilmu pengetahuan dan teknologi; perusahaan-perusahaan besar seperti Google, Microsoft, dan IBM berkomitmen pada AI dan mengaplikasikannya dalam berbagai bidang. AI adalah teknologi multidisiplin yang memiliki kemampuan untuk mengintegrasikan kognisi, pembelajaran mesin, pengenalan emosi, interaksi manusia-komputer, penyimpanan data, dan pengambilan keputusan. Pertama kali diusulkan oleh John McCarthy dalam Konferensi Dartmouth pada pertengahan abad ke-20, AI telah mencapai beberapa hasil yang signifikan sejak tahun 1993. Berkat

penggunaan luas algoritma BP, jaringan saraf telah berkembang dengan cepat. Di lingkungan yang luas, penggunaan sistem pakar telah menghemat biaya industri dan meningkatkan efisiensi industri. Misalnya, sistem pakar PROSPECTOR berhasil menganalisis deposit mineral bernilai ratusan juta dolar. Setelah itu, orang mulai mencoba untuk mempelajari program kecerdasan buatan yang umum, tetapi menghadapi hambatan serius dan mencapai jalan buntu. Kecerdasan buatan kembali berada pada titik rendah. Pada tahun 1997, kesuksesan "Deep Blue" membawa pengembangan AI ke dalam agenda. Dengan peningkatan daya komputasi, bottleneck AI berhasil dipecahkan, dan pengembangan pembelajaran dalam mendalam dan pembelajaran berbasis data besar terus berlanjut. Dengan perkembangan terus-menerus dari GPU, pengembangan penuh kekuatan komputasi telah terus meningkat; ini telah membentuk dasar bagi perkembangan pesat AI. Kecerdasan buatan telah mengalami proses pengembangan yang panjang, dengan sejarah lebih dari 70 tahun. Proses pengembangan ini dapat dibagi menjadi beberapa tahap: pada tahun 1943, model neuron buatan pertama kali diusulkan, yang membuka era penelitian jaringan saraf buatan. Pada tahun 1956, Konferensi Dartmouth diadakan dan konsep kecerdasan buatan diperkenalkan; ini menandai lahirnya kecerdasan buatan. Selama periode ini, tren penelitian kecerdasan buatan oleh komunitas akademik internasional mengalami peningkatan, dan pertukaran ilmiah berlangsung sering. Pada tahun 1960-an, jenis utama dari konneksionisme dan tunduk pada kehilangan popularitas, dan teknologi pintar mengalami

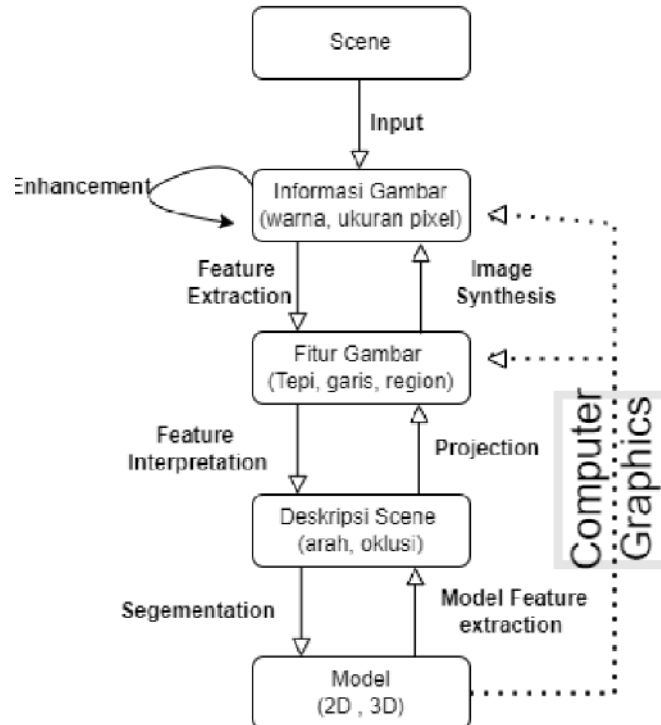
penurunan dalam pengembangannya. Penelitian tentang algoritma backpropagation dimulai pada tahun 1970-an, dan biaya dan daya komputasi komputer secara bertahap meningkat, sehingga membuat penelitian dan aplikasi sistem pakar menjadi sulit. Melanjutkan menjadi sulit, tetapi kecerdasan buatan secara bertahap membuat terobosan. Pada tahun 1980-an, jaringan saraf berdasarkan backpropagation menjadi dikenal luas, penelitian algoritma berdasarkan jaringan saraf buatan berkembang dengan pesat, fungsi perangkat keras komputer telah berkembang pesat, dan perkembangan Internet mengurangi perkembangan kecerdasan buatan. Di dekade pertama abad ke-21, perkembangan Internet mobile membawa lebih banyak skenario aplikasi kecerdasan buatan. Pada tahun 2012, pembelajaran mendalam diusulkan, dan kecerdasan buatan mencapai perkembangan terobosan; algoritma ini mencapai terobosan dalam pengenalan suara dan visual (Zhang dkk, 2021).

2.1.2. Computer Vision

Computer vision (CV) atau visi komputer dikembangkan dengan memiliki dua tujuan ganda. Dari sudut ilmu sains biologi, visi komputer dimaksudkan untuk memahami model komputasi dari sistem visual manusia. Dari sudut ilmu rekayasa titik pandang, CV bertujuan membangun sistem otonom yang dapat melakukan beberapa tugas yang dapat dilakukan oleh sistem visual manusia (bahkan diharapkan dapat melampauinya dalam beberapa kasus). Banyaknya tantangan di bidang ini terkait dengan ekstraksi informasi 3D dan temporal dari

dari 2D dengan waktu pengambilan yang variatif sebagaimana satu atau lebih kamera memperoleh informasi yang terkandung dalam adegan dinamis (Huang, n.d.). Tentu saja, kedua tujuan tersebut berkaitan erat. Keberagaman mata manusia dan sifat sistem visual manusia sering memberikan inspirasi kepada para insinyur yang merancang sistem visi komputer. Sebaliknya, algoritma visi komputer dapat menawarkan wawasan tentang bagaimana sistem visual manusia bekerja dalam sekian baris kode.

Pemrosesan CV pada lingkup objek 3D terdiri dari banyak komponen yang melakukan fungsi tertentu dalam koordinasi dengan komponen lain. Gambar 2.1 merupakan blok rangkaian utama dan tipe informasi atau data yang terlibat dari proses tersebut (Shirai, 1987).

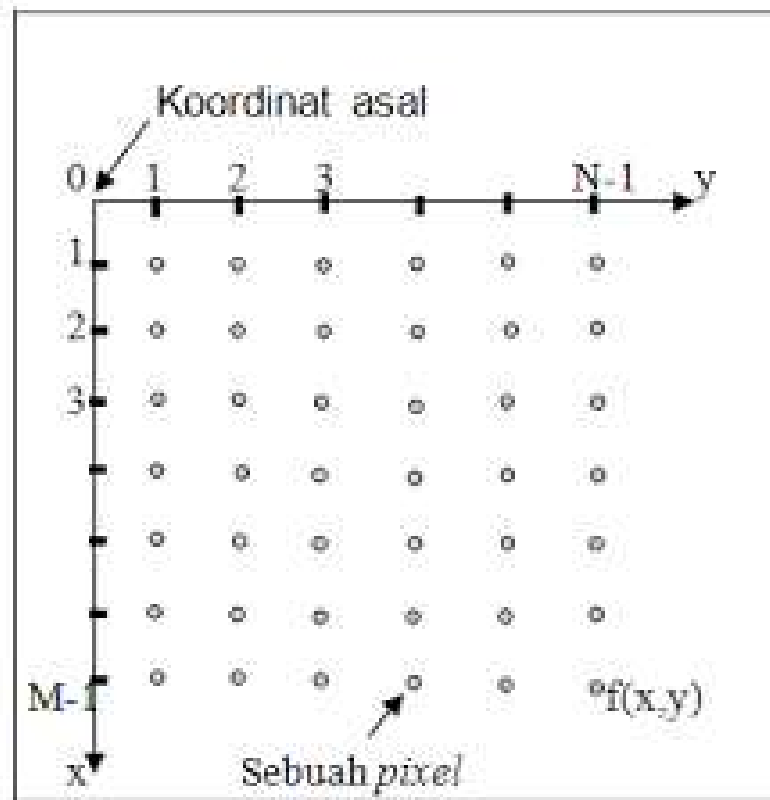


Gambar 2. 1 Rangkaian umum proses computer vision

Pemrosesan CV pada lingkup objek 3D terdiri dari banyak komponen yang melakukan fungsi. Sebagaimana yang tersaji pada gambar 2.1, Ketika sebuah adegan/scene masuk ke dalam program, informasi dasar dari gambar tersebut akan diekstraksi baik komponen warna pada gambar, ukuran tiap objek dalam pixel serta informasi pixel keseluruhan adegan itu sendiri. Informasi yang diekstraksi bisa berupa raw gambar yang original, maupun scene yang telah mengalami beberapa enhancement baik itu kontras, saturasi, dll. Setelah itu scene tersebut akan diinterpolasi untuk mendapatkan data tepi, garis, region, dan data apapun yang berkaitan dengan pembentukan dasar tiap objek pada scene tersebut. Langkah terakhir merupakan segmentasi untuk mendapatkan deskripsi mengenai scene tersebut yang selanjutnya akan terbentuk matriks model 2D maupun 3D dari satu scene tersebut.

2.1.3. Pengolahan Citra Digital

Pengolahan citra digital adalah pemrosesan gambar dua dimensi menggunakan komputer dan mengacu pada pemrosesan setiap data dua dimensi. Citra digital merupakan sebuah larik yang berisi nilai nilai real maupun kompleks yang direpresentasikan dengan deretan bit tertentu. Suatu citra data didefinisikan sebagai fungsi $f(x, y)$ berukuran M untuk baris dan N untuk kolom, dimana x dan y adalah koordinat spasial dan amplitudo f di titik koordinat (x, y) dinamakan intensitas atau tingkat keabuan dari citra pada titik tersebut adalah citra digital, sebagaimana yang ditunjukkan pada gambar 2.1 berikut:



Gambar 2. 2 Koordinat Citra Digital

Sebuah citra digital dapat diwakili oleh sebuah matriks dua dimensi $f(x, y)$ yang terdiri dari M kolom dan N baris, dimana perpotongan antara kolom dan baris disebut piksel (pixel = *picture element*) atau elemen terkecil dari sebuah citra. Adapun bentuk bentuk matriks $f(x, y)$, sebagaimana yang ditunjukkan persamaan (1) berikut ini.

$$\begin{aligned}
 f(x, y) [& f(0,0) & f(0,1) & \cdots & f(0, N-1) \\
 & f(1,0) & f(1,1) & \cdots & f(1, N-1) \\
 & \vdots & \vdots & & \vdots \\
 & f(M-1,0) & f(M-1,1) & \cdots & f(M-1, N-1)] \dots \dots \dots (1)
 \end{aligned}$$

Suatu citra $f(x, y)$ dalam fungsi matematis dapat dituliskan sebagai berikut:

$$0 \leq x \leq M - 1$$

$$0 \leq y \leq N - 1$$

$$0 \leq f(x, y) \leq G - 1$$

dimana:

M = jumlah piksel baris (*row*) pada array citra

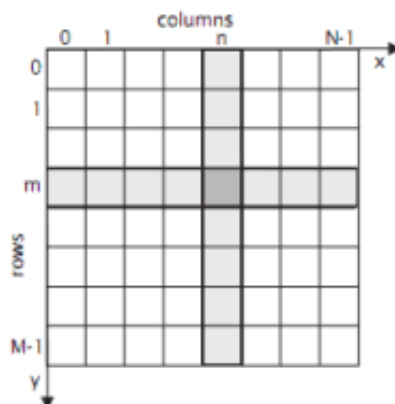
N = jumlah piksel kolom (*column*) pada array citra

G = Nilai skala keabuan (*grayscale*)

Besarnya nilai M , N dan G pada umumnya merupakan perpangkatan dari dua.

$$M = 2^m; N = 2^n; G = 2^k \dots\dots\dots (2)$$

dimana nilai m , n dan k adalah bilangan bulat positif. Interval $(0, G)$ disebut skala keabuan (*grayscale*). Besar G tergantung pada proses digitalisasinya. Biasanya keabuan 0 (nol) menyatakan intensitas hitam dan 1 (satu) menyatakan intensitas putih. Untuk citra 8 bit, nilai G sama dengan $2^8 = 256$ warna (derajat keabuan) (Putra, 2010) (Kusumanto & Tompunu, 2016).



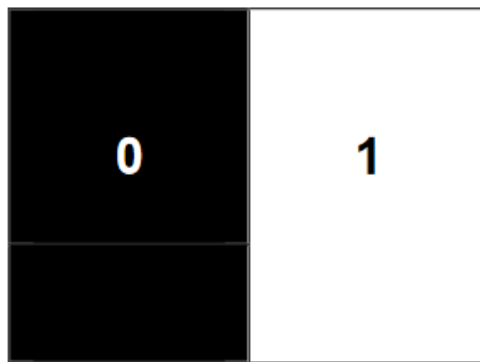
Gambar 2. 3. Representasi citra digital dalam dua dimensi

2.1.4. Jenis Citra Digital

Ada banyak cara untuk menyimpan citra digital di dalam memori, cara penyimpanan menentukan jenis citra digital yang terbentuk. Beberapa jenis citra digital yang sering digunakan adalah citra biner, citra skala keabuan, dan citra warna (T. Sutoyo, S.Si, Edy Mulyanto, S.Si., 2009).

A. Citra Biner

Banyaknya warna yakni dua, yaitu hitam dan putih, dibutuhkan 1 bit di memori untuk menyimpan kedua warna ini, dimana gradasi warnanya adalah bit 0 untuk warna hitam dan bit 1 untuk warna putih, seperti yang ditunjukkan pada gambar 2.4.



Gambar 2. 4. *Citra Biner Dua Warna*

B. Citra Skala Keabuan (*Greyscale*)

Banyaknya warna tergantung pada jumlah bit yang disediakan di memori untuk menampung kebutuhan warna. Citra dua bit mewakili 4 warna dengan gradasi warna sebagaimana yang ditunjukkan oleh

gambar 2.5, sedangkan untuk Citra tiga bit, mewakili 8 warna dengan gradasi warna sebagaimana terlihat pada gambar 2.6.

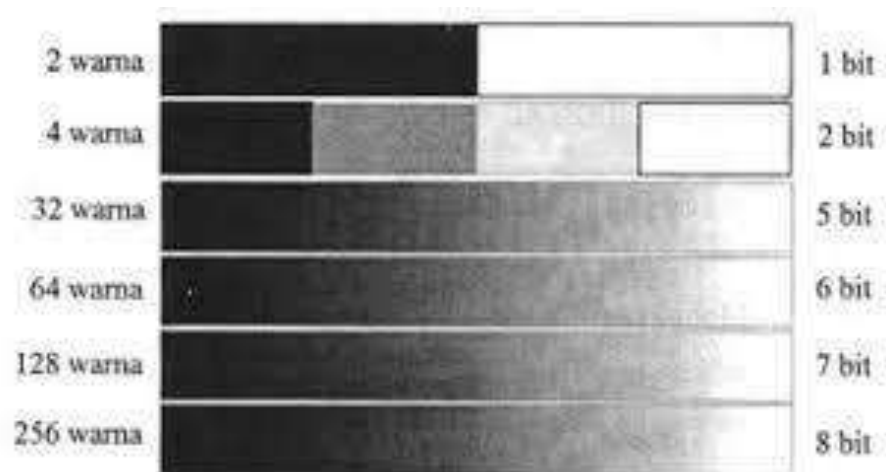


Gambar 2. 5. Citra dua bit empat warna



Gambar 2. 6. Citra tiga delapan warna

Semakin besar jumlah bit warna yang disediakan di memori, maka semakin halus gradasi warna yang terbentuk. Gambar 2.7 menunjukkan perbandingan gradasi warna untuk jumlah bit mulai dari 2 bit, sampai dengan 8 bit.



Gambar 2. 7. Citra tiga delapan warna

C. Citra Warna

Setiap piksel pada citra warna mewakili warna yang merupakan kombinasi dari tiga warna dasar (RGB = Red, Green, Blue). Setiap warna dasar menggunakan penyimpanan 8 bit = 1 byte, yang berarti setiap warna mempunyai gradasi sebanyak 255 warna. Berarti setiap piksel mempunyai kombinasi warna sebanyak $2^8 \cdot 2^8 \cdot 2^8 = 224\ 16$ juta warna lebih. Itulah sebabnya format ini dinamakan true color karena mempunyai jumlah warna yang cukup besar sehingga bisa dikatakan hamper mencakup semua warna di alam.

Penyimpanan citra warna di dalam memori berbeda dengan citra skala keabuan. Setiap piksel dari citra skala keabuan 256 gradasi warna diwakili oleh 1 byte. Sedangkan 1 piksel citra warna diwakili oleh 3 byte, dimana masing-masing byte merepresentasikan warna merah, hijau, dan biru, gambar 2.8 adalah contoh penyimpanan citra warna di dalam memori.

Yellow R = 255 G = 255 B = 0	Orange R = 255 G = 102 B = 0	Green R = 0 G = 255 B = 0
Cyan R = 0 G = 255 B = 255	Violet R = 204 G = 102 B = 204	White R = 255 G = 255 B = 255
Black R = 0 G = 0 B = 0	Turquoise R = 102 G = 255 B = 204	Brown R = 153 G = 102 B = 51

Gambar 2. 8. *Penyimpanan Citra Warna di dalam memori*

2.1.5. Elemen Citra Digital

Berikut ini merupakan elemen-elemen yang terdapat pada citra digital (Putra, 2010):

a. Kecerahan (*Brightness*)

Kecerahan merupakan intensitas cahaya yang dipancarkan piksel dari citra yang dapat ditangkap oleh sistem penglihatan. Kecerahan pada sebuah titik (piksel) di dalam citra merupakan intensitas rata-rata dari suatu area yang melingkupinya.

b. Kontras

Kontras menyatakan sebaran terang dan gelap dalam sebuah citra. Pada citra yang baik, komposisi gelap dan terang tersebar secara merata

c. Kontur

Kontur adalah keadaan yang ditimbulkan oleh perubahan intensitas pada piksel –piksel yang bertetangga. Karena adanya perubahan intensitas, maka mata mampu mendeteksi tepi-tepi objek di dalam citra.

d. Warna

Warna sebagai persepsi yang ditangkap sistem visual terhadap panjang gelombang cahaya yang ditentukan oleh objek.

e. Bentuk

Bentuk adalah property intrinsik dari objek 3 dimensi, dengan pengertian bahwa bentuk merupakan properti intrinsik utama untuk sistem visual manusia.

f. Tekstur

Tekstur dicirikan sebagai distribusi spasial dari derajat keabuan di dalam sekumpulan piksel-piksel yang bertetangga. Tekstur adalah sifat-sifat atau karakteristik yang dimiliki oleh suatu daerah yang cukup besar sehingga secara alami sifat-sifat tadi dapat berulang dalam daerah tersebut. Tekstur adalah keteraturan pola-pola tertentu yang terbentuk dari susunan piksel-piksel dalam citra digital. Informasi tekstur dapat digunakan untuk membedakan sifat-sifat permukaan suatu benda dalam citra yang berhubungan dengan kasar dan halus, juga sifat-sifat spesifik dari kekasaran dan kehalusan permukaan tadi, yang sama sekali terlepas dari warna permukaan tersebut.

2.1.6. Resolusi Citra Digital

Resolusi citra adalah tingkat detail dari suatu citra, semakin tinggi resolusi sebuah citra digital, maka semakin tinggi tingkat detail dari citra tersebut. Resolusi merupakan jumlah piksel yang tersusun dalam sebuah gambar digital. Resolusi berbanding lurus dengan kualitas gambar, karena itu resolusi juga merupakan salah satu faktor penentu kualitas dari citra digital. Semakin tinggi resolusi, semakin baik kualitas gambar digital, dan sebaliknya jika resolusinya semakin rendah maka semakin rendah kualitas gambar.

Setiap jenis citra memiliki keunggulan masing-masing dalam hal resolusi, suatu citra dapat unggul untuk resolusi tertentu namun lemah untuk jenis resolusi yang lain. Terdapat dua jenis resolusi citra digital yang perlu diketahui (T. Sutoyo, S.Si, Edy Mulyanto, S.Si., 2009), yaitu:

a. Resolusi Spasial

Resolusi spasial ini merupakan ukuran halus (resolusi tinggi) atau kasar (resolusi rendah) pada pembagian kisi-kisi baris dan kolom pada saat sampling, resolusi ini dipakai untuk menentukan jumlah pixel per satuan panjang. Biasanya satuan resolusi ini adalah dpi (dot per inchi), resolusi ini sangat berpengaruh pada detail dan perhitungan gambar.

b. Resolusi Kecemerlangan

Resolusi kecermerlangan (intensitas/ brightness) atau biasanya disebut dengan kedalaman bit/kedalaman warna (Bit Depth) adalah ukuran halus (resolusi tinggi) atau kasar (resolusi rendah) pembagian tingkat gradasi warna saat dilakukan kuantisasi. Bit Depth menentukan berapa banyak informasi warna yang tersedia untuk ditampilkan dalam setiap piksel. Semakin besar nilainya, semakin bagus kualitas gambar yang dihasilkan dan tentu ukuran juga semakin besar.

Pada sistem pengenalan wajah yang telah berkembang sejak tahun 1960-an, untuk mengidentifikasi seseorang dari gambar digital atau video, citra atau gambar diklasifikasi menjadi dua bagian yaitu gambar dengan resolusi tinggi (halus) dan gambar dengan resolusi rendah (kasar). Adapun ukuran gambar yang dikategorikan sebagai gambar dengan resolusi rendah yakni 64 x 64 pixels, 32 x 32 pixels dan 16 x 16 pixels. Gambar dengan kategori beresolusi tinggi dimulai dari ukuran 128 x 128 pixels (Engineering & Gables, 2017).

2.1.7. Video Digital

Video adalah sekumpulan citra yang direkam atau diakuisisi selama satu satuan waktu tertentu. Citra di dalam video lebih dikenal dengan istilah frame atau frame citra. Kecepatan standar perekaman dan visualisasi video disesuaikan dengan kecepatan mata manusia dalam memahami dan menganalisis informasi yang terkandung di dalam setiap frame. Pada kamera video standar Eropa kecepatannya adalah 25 frame per detik/frame per second (fps), sedangkan untuk standar Amerika adalah 30 fps. Jika sekumpulan citra ditampilkan ke layar monitor dengan kecepatan tampil antara 25 hingga 30 fps maka mata manusia dapat mengamati dengan baik informasi dan pergerakan objek di dalam setiap frame video secara kontinu. Bila tampilan frame video ini diperlambat menjadi di bawah 15 fps maka akan tampak pergerakan objek yang tidak kontinyu (terputus-putus), sebaliknya bila tampilan frame video ini dipercepat di atas 40 fps maka pergerakan objek akan semakin cepat dan mata manusia semakin sulit untuk menganalisis informasi pergerakan objek (Madenda, 2015).

Video terbagi dalam dua kategori yaitu video analog dan video digital. Video analog mengkodekan informasi gambar dengan memvariasikan voltase dan/atau frekuensi dari sinyal. Seluruh sistem sebelum video digital dapat dikategorikan sebagai video analog. Sementara video digital dapat disebut array 3 dimensi dari pixel berwarna, 2 dimensi melayani arah spesial dari gambar bergerak (horizontal dan vertikal) dan satu dimensinya akan mempresentasikan domain waktu (Wibowo et al., 2017).

2.1.8. Wajah

Berdasarkan KBBI (Kamus Besar Bahasa Indonesia) secara etimologi wajah adalah bagian depan dari kepala; roman muka; muka. Wajah pada manusia menjadi fokus perhatian di dalam interaksi sosial, selain itu wajah memiliki peran penting dalam mengidentifikasi seseorang. Adapun batasan wajah pada manusia meliputi wilayah antara dahi dan dagu termasuk rambut, alis, pelipis, mata, hidung, pipi, mulut, bibir, gigi, kulit, dahi dan dagu (Badan Pengembangan dan Pembinaan Bahasa, 2016b).

Tujuan pengembangan aplikasi biometrik, seperti pengenalan wajah, telah menjadi penting dalam konteks kota cerdas. Selain itu, banyak ilmuwan dan insinyur di seluruh dunia telah berfokus pada pembentukan algoritma dan metode yang semakin kuat dan akurat untuk jenis sistem ini dan penerapannya dalam kehidupan sehari-hari. Semua jenis sistem keamanan harus melindungi semua data pribadi. Jenis yang paling umum digunakan untuk pengenalan adalah kata sandi. Namun, melalui perkembangan teknologi informasi dan algoritma keamanan, banyak sistem mulai menggunakan banyak faktor biometrik untuk tugas pengenalan. Faktor-faktor biometrik ini memungkinkan identifikasi identitas seseorang berdasarkan karakteristik fisiologis atau perilaku mereka. Mereka juga memberikan beberapa keuntungan, misalnya, kehadiran seseorang di depan sensor sudah cukup, dan tidak perlu lagi mengingat beberapa kata sandi atau kode rahasia. Dalam konteks ini, banyak sistem pengenalan berdasarkan berbagai faktor biometrik seperti iris, sidik jari, suara, dan wajah telah diterapkan dalam beberapa tahun terakhir.

Sistem yang mengidentifikasi individu berdasarkan karakteristik biologis mereka sangat menarik karena mudah digunakan. Wajah manusia terdiri dari berbagai struktur dan karakteristik. Karena itu, dalam beberapa tahun terakhir, pengenalan wajah telah menjadi salah satu sistem otentikasi biometrik yang paling banyak digunakan, mengingat potensinya dalam banyak aplikasi dan bidang (pengawasan, keamanan rumah, kontrol perbatasan, dan sebagainya). Sistem pengenalan wajah sebagai ID (identitas) sudah ditawarkan kepada konsumen di luar telepon seluler, termasuk pada check-in bandara, stadion olahraga, dan konser. Selain itu, sistem ini tidak memerlukan campur tangan manusia untuk beroperasi, yang memungkinkan pengidentifikasian individu hanya dari gambar yang diperoleh dari kamera. Selain itu, banyak sistem biometrik yang dikembangkan menggunakan berbagai jenis pencarian memberikan akurasi identifikasi yang baik. Namun, akan menarik untuk mengembangkan sistem biometrik baru untuk pengenalan wajah guna mencapai batasan waktu nyata (Kortli, 2020).

2.1.9. Face Detection

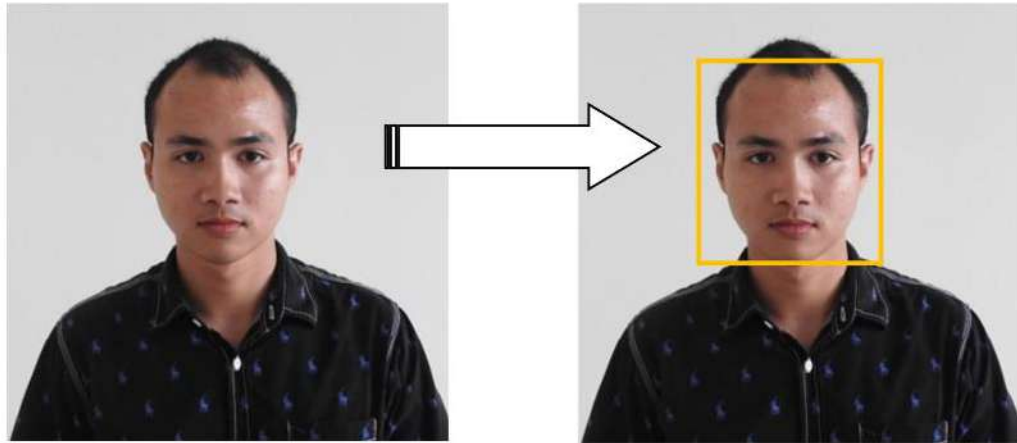
Deteksi wajah adalah teknologi komputer yang digunakan dalam berbagai aplikasi yang mengidentifikasi lokasi adanya wajah manusia dalam gambar digital. Deteksi wajah juga merupakan salah satu tahap pra-proses yang sangat penting di dalam sistem pengenalan wajah (face recognition) yang digunakan untuk sistem biometric (Viola & Jones, 2001) . Deteksi wajah juga dapat digunakan untuk pencarian dan pengindeksan citra atau video yang di dalamnya terdapat wajah manusia dalam berbagai ukuran, posisi, dan latar

belakang. Pemrosesan gambar ini bertujuan untuk mencari wajah dari gambar obyek bergerak yang telah di-capture, kemudian gambar tersebut diolah dengan memisahkan gambar dengan latar belakangnya, sehingga hanya bagian yang dianggap kulit yang ditampilkan sedangkan bagian yang bukan kulit akan dihitamkan. Pemisahan gambar dengan latar belakang ini bertujuan untuk memudahkan proses pencarian wajah.

Deteksi wajah adalah menentukan keberadaan wajah dalam gambar dan jika ada, mencari lokasi gambar dan bagian masing-masing wajah. Deteksi wajah merupakan hal yang mudah bagi manusia, namun sangat sulit untuk komputer (Yang et al., 2016). Tujuan utama dari deteksi wajah adalah mengetahui semua posisi wajah pada gambar dan memberi tanda dengan kotak persegi (Ma & Wang, 2019).

Deteksi wajah adalah langkah awal untuk melakukan identifikasi wajah, analisis wajah, estimasi pose wajah dan lainnya. Sebuah pendeteksi wajah yang ideal seharusnya mampu mengidentifikasi dan menemukan lokasi dan luas semua wajah yang ada di dalam sebuah gambar tanpa memperhatikan pose, skala, orientasi, umur, dan ekspresi (Li et al., 2004). Tingkat pemrosesan yang lebih tinggi dapat dilakukan hanya jika wajah ditandai secara akurat dalam gambar atau video dan landmark wajah secara akurat berada. Proses pendeteksi wajah ini bekerja dengan cara memeriksa citra yang dimasukkan, apakah memiliki citra wajah atau tidak, jika memiliki, maka akan dilakukan pemisahan dengan cara memotong citra wajah dari latar belakang citra yang dimasukkan. Berikut merupakan gambar

pendeteksian wajah. Gambar 2.2 berikut merupakan contoh hasil deteksi wajah



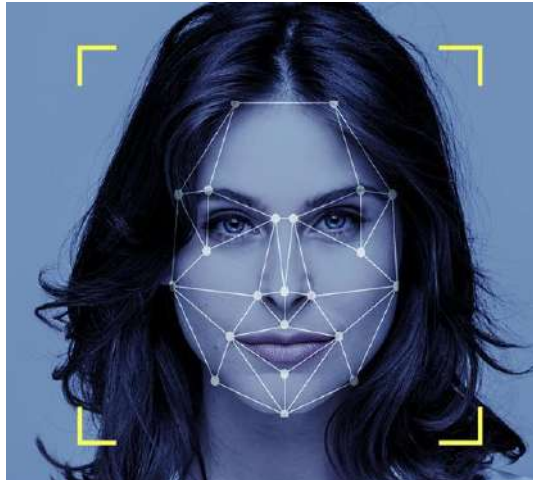
Gambar 2. 9 Ilustrasi Hasil deteksi wajah

2.1.10. Face Recognition

Pengenalan Wajah atau face recognition adalah salah satu teknik biometric yang memungkinkan komputer untuk mengenali wajah manusia yang berorientasi pada wajah yang telah diterapkan untuk banyak aplikasi dalam bidang keamanan, antara lain Access security system, Authentication system, hingga sebagai alat bantu dalam pelacakan pelaku kriminal. Pengenalan ini dapat dibagi menjadi dua bagian yaitu: “dikenali” dan “tidak dikenali”, setelah dilakukan dengan pola yang sebelumnya disimpan di dalam database. Sistem ini juga harus bisa mengenali objek bukan wajah.

Secara umum sistem pengenalan citra wajah dibagi menjadi dua jenis, yaitu sistem feature-based dan sistem image-based. Pada sistem pertama digunakan ciri yang dieksekusi dari komponen citra wajah seperti

mata, hidung, mulut, dan lain-lain yang kemudian dimodelkan secara geometris hubungan antara ciri-ciri tersebut. Sedangkan pada sistem kedua menggunakan informasi dari pixel citra yang kemudian direpresentasikan dalam metode tertentu, yang kemudian digunakan untuk klasifikasi identitas citra (Fatta, 2009). Hasil dari face recognition adalah pemetaan fitur pada wajah yang kemudian dicocokkan dengan database. Gambar 2.3 berikut merupakan contoh hasil face recognition (MIT News, 2022).



Gambar 2. 10 Contoh hasil rekognisi wajah

Pengenalan wajah (Face Recognition, FR) telah menjadi bidang penelitian yang aktif dalam beberapa tahun terakhir karena berbagai aplikasi yang dapat diterapkannya, seperti keamanan perbatasan, pengawasan, penegakan hukum, dan pengendalian akses. Baru-baru ini, aplikasi lain yang terkait dengan sistem FR meliputi grafika komputer, jaringan saraf, dan psikologi karena semakin menjadi minat multidisiplin.

Seperti sistem biometrik lainnya, seperti yang ditunjukkan dalam Gambar 2.3, tahapan yang terlibat dalam proses FR meliputi deteksi wajah, pra-pemrosesan citra wajah, ekstraksi fitur wajah, dan klasifikasi fitur terakhir. Tahap pertama, yaitu deteksi wajah, adalah proses di mana sistem memverifikasi keberadaan wajah dalam gambar atau video. Setelah wajah terdeteksi, citra pra-diproses untuk memperoleh wilayah minat dan meningkatkan kualitas gambar. Normalisasi adalah jenis teknik pra-pemrosesan di mana gambar wajah dengan skala yang berbeda diubah dan dipetakan ke skala yang sama. Penyelarasan wajah adalah pendekatan pra-pemrosesan lain yang merupakan proses lokal mengidentifikasi titik fidusial seperti mulut, mata, dagu, dan hidung. Pendekatan ini terbukti meningkatkan sistem pengenalan wajah, meskipun tetap menjadi tantangan dan masalah terbuka dalam lingkungan tanpa batasan. Peningkatan citra juga merupakan pendekatan pra-pemrosesan yang sering diabaikan dalam literatur. Tujuan utamanya dalam hal ini adalah menghasilkan citra wajah yang ditingkatkan dari citra asli, yang seharusnya meningkatkan kinerja keseluruhan sistem pengenalan wajah.

Ekstraksi fitur adalah tahap berikutnya dari model sistem pengenalan wajah, dan tujuannya adalah menyederhanakan jumlah sumber daya yang menggambarkan kumpulan data besar. Selain itu, ekstraksi fitur unik dilakukan untuk meminimalkan noise dan informasi yang tidak relevan yang ada dalam citra wajah asli, dan vektor fitur yang cukup untuk menggambarkan wajah dari citra wajah diekstraksi. Banyak metode ekstraksi fitur telah diusulkan dalam literatur, namun memilih fitur yang tepat untuk berbagai sistem pengenalan

wajah dalam lingkungan tanpa batasan tetap menjadi tantangan. Klasifier paling umum yang digunakan dalam bidang pengenalan wajah adalah classifier jarak minimum, classifier tetangga terdekat, dan classifier k-tetangga terdekat. Classifier jarak minimum menempatkan label sampel uji sebagai kelas yang rata-ratanya terkait. Classifier tetangga terdekat menempatkan sampel uji ke dalam kelas yang terdekat tetangganya, sedangkan classifier k-tetangga terdekat menempatkan sampel uji ke dalam kelas yang memiliki tetangga terdekat dengan mencari terlebih dahulu k-tetangga terdekat.

Dalam beberapa tahun terakhir, algoritma pembelajaran mesin telah menjadi yang paling disukai untuk menggunakan teknik klasifikasi seperti jaringan saraf konvolusi (CNN). Tahap klasifikasi fitur yang mengarah ke pengenalan citra wajah melibatkan identifikasi dan otentikasi citra wajah. Identifikasi membandingkan citra wajah dengan citra wajah lain untuk dapat mengidentifikasi wajah di antara beberapa kemungkinan, sedangkan otentikasi adalah ketika sebuah wajah dibandingkan dengan wajah lain untuk menyetujui identitas yang diminta. Dalam kedua skenario tersebut, citra wajah individu yang dikenal terdaftar dalam sistem yang dikenal sebagai galeri. Setelah itu, citra wajah yang digunakan sebagai probe dapat menjadi individu yang terdaftar atau tidak terdaftar yang digunakan untuk tugas identifikasi atau pengenalan (Adjabi dkk, 2020).

2.1.11. Object Tracking

Objek Tracking merupakan teknik yang digunakan dalam bidang komputer vision untuk mendeteksi dan melacak objek yang bergerak dalam suatu rangkaian video atau gambar secara real-time (Yilmaz et al., 2006). Teknik ini telah menjadi area penelitian yang sangat penting dalam pengembangan sistem pengenalan objek, pengawasan keamanan, navigasi kendaraan otonom, augmented reality, dan banyak lagi.

Pendekatan yang umum digunakan dalam object tracking adalah pendekatan berbasis deteksi (Yilmaz et al., 2006). Pendekatan ini terdiri dari dua tahap utama, yaitu tahap deteksi objek pada frame awal dan tahap pelacakan objek pada frame-frame berikutnya. Pada tahap deteksi, algoritma deteksi digunakan untuk mengenali objek yang akan dilacak. Deteksi dapat dilakukan dengan menggunakan teknik-teknik seperti deteksi wajah, deteksi manusia, atau deteksi objek lainnya.

Setelah objek berhasil dideteksi pada frame awal, tahap pelacakan dimulai. Terdapat berbagai algoritma pelacakan yang umum digunakan, seperti pelacakan berbasis titik-fitur, pelacakan berbasis histogram, pelacakan berbasis filter Kalman, dan pelacakan berbasis Deep Learning (Yilmaz et al., 2006; Khan et al., 2018). Setiap algoritma pelacakan memiliki kelebihan dan keterbatasan sendiri, dan pemilihan algoritma yang tepat tergantung pada kebutuhan aplikasi serta karakteristik objek yang akan dilacak.

Salah satu tantangan utama dalam object tracking adalah ketika objek mengalami perubahan skala, rotasi, atau occlusion. Objek juga dapat berpindah ke latar belakang yang memiliki warna atau tekstur yang mirip, sehingga

mempersulit pelacakan (Yilmaz et al., 2006). Untuk mengatasi tantangan ini, penelitian terus dilakukan untuk mengembangkan algoritma pelacakan yang lebih robust dan akurat.

2.1.12. Oklusi Pengenalan Wajah

Oklusi berasal dari kata occlusion, yang terdiri dari dua kata yakni oc yang berarti ke atas (up) dan clusion yang berarti menutup (closing). Jadi oklusi pada pengenalan wajah merupakan wajah – wajah yang tertangkap oleh kamera yang dapat tertutup sebagian dari wajahnya antara lain oleh aksesoris, janggut, atau kumis. Oklusi pada pengenalan wajah adalah tantangan untuk setiap peneliti bagaimana cara mengurangi halangan pada proses pengenalan wajah yang akan dilakukan oleh sistem (Yang et al., 2016).

2.1.13. Surveillance Detection System

Surveillance Detection System (SDS) adalah teknologi canggih yang digunakan dalam sistem keamanan untuk mengidentifikasi dan mengatasi aktivitas pengawasan potensial. SDS memainkan peran penting dalam meningkatkan langkah-langkah keamanan di lokasi sensitif seperti bandara, fasilitas pemerintah, dan infrastruktur kritis. SDS menggunakan kombinasi sensor, kamera, dan algoritma pintar untuk memantau dan menganalisis lingkungan sekitar terhadap tanda-tanda pengawasan yang mencurigakan (Ouahada, Koller, & Spiekermann, 2016). Sistem ini dirancang untuk mendeteksi pola perilaku mencurigakan, gerakan yang tidak biasa, atau individu

yang tidak diizinkan yang dapat membahayakan keamanan area yang dilindungi.

SDS bekerja dengan terus-menerus mengumpulkan data dari berbagai sumber seperti kamera pengawasan video, detektor gerakan, sistem kontrol akses, dan aliran media sosial (Djeziri, 2018). Data yang dikumpulkan diproses dan dianalisis menggunakan algoritma pembelajaran mesin untuk mengidentifikasi aktivitas atau pola yang tidak normal yang dapat mengindikasikan pengawasan. Algoritma-algoritma ini dilatih untuk membedakan antara perilaku normal dan perilaku pengawasan potensial berdasarkan aturan dan model yang telah ditentukan sebelumnya. Begitu ancaman potensial terdeteksi, sistem menghasilkan peringatan secara real-time kepada personel keamanan, memungkinkan mereka mengambil tindakan yang tepat untuk mengurangi risiko (Petitcolas, 2019).

Keefektifan SDS bergantung pada integrasi teknologi-teknologi yang beragam, termasuk visi komputer, pengenalan pola, analisis data, dan kecerdasan buatan (Ouahada, Koller, & Spiekermann, 2016). Dengan menggabungkan teknologi-teknologi ini, sistem dapat secara akurat mengidentifikasi perilaku mencurigakan dan memberikan intelijen yang dapat dijadikan dasar tindakan kepada tim keamanan. Selain itu, SDS dapat ditingkatkan melalui integrasi sistem keamanan lainnya, seperti pengenalan wajah dan pengenalan nomor plat, untuk memperkuat kemampuannya dalam mengidentifikasi ancaman potensial (Chen, Chiang, & Su, 2020).

Penerapan SDS semakin penting dalam lanskap keamanan yang terus berkembang saat ini. Dengan meningkatnya aktivitas teroris dan perlunya melindungi infrastruktur kritis, terdapat permintaan yang semakin meningkat untuk teknologi pengawasan yang canggih (Djeziri, 2018). SDS tidak hanya bertindak sebagai penghalang terhadap ancaman potensial, tetapi juga memungkinkan personel keamanan untuk merespons dengan cepat dan efektif terhadap aktivitas mencurigakan. Selain itu, penelitian dan pengembangan yang terus dilakukan dalam bidang SDS bertujuan untuk meningkatkan akurasi, skalabilitas, dan adaptabilitasnya dalam berbagai lingkungan, menjadikannya komponen penting dari sistem keamanan modern (Petitcolas, 2019).

2.1.14. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN), atau Jaringan Saraf Konvolusi dalam bahasa Indonesia, adalah salah satu jenis arsitektur jaringan saraf tiruan yang sangat efektif dalam pengolahan data visual, seperti gambar dan video. Arsitektur CNN terinspirasi oleh fungsi korteks visual manusia. Ini dirancang untuk menyerupai hubungan antara neuron dalam otak manusia. Pengenalan gambar adalah tugas yang telah dilakukan oleh manusia sejak kecil. Anak-anak diajarkan untuk mengidentifikasi buah-buahan dan sayuran, seperti apel, pisang, dan semangka. Apakah mungkin untuk mengajarkan komputer melakukan hal yang sama? Apakah mungkin bagi manusia membangun mesin yang bisa melihat dan memahami seperti manusia? Jawabannya adalah ya untuk semua pertanyaan ini. Manusia harus menunjukkan algoritma dengan jutaan gambar sebelum komputer dapat menggeneralisasi input dan membuat prediksi untuk

gambar yang belum pernah dilihat sebelumnya, persis seperti manusia harus menunjukkan algoritma dengan jutaan gambar sebelum komputer dapat mengeneralisasi input dan membuat prediksi untuk gambar yang belum pernah dilihat sebelumnya. Pertanyaannya adalah: bagaimana gambar terlihat bagi komputer? Gambar 1 Manusia dapat melihat gambar angka 8, tetapi komputer dapat melihat data numerik. Oleh karena itu, mengembangkan komputer yang dapat menganalisis dan mengenali gambar adalah tugas yang kompleks (Bhatt dkk, 2021).



Gambar 2. 11 Bagaimana Komputer melihat gambar

A. Input Image

Pixel merupakan unsur dasar dalam suatu gambar komputer. Mereka mewakili data visual dalam bentuk biner. Kumpulan pixel yang memiliki nilai dari 0 hingga 255 disusun dalam susunan mirip matriks pada gambar digital. Setiap nilai pixel mengindikasikan kecerahan dan nuansa dari masing-masing pixel tersebut. Ketika manusia mengamati sebuah gambar, otak manusia mampu memproses sejumlah besar informasi dalam sepersekian detik pertama. Dalam

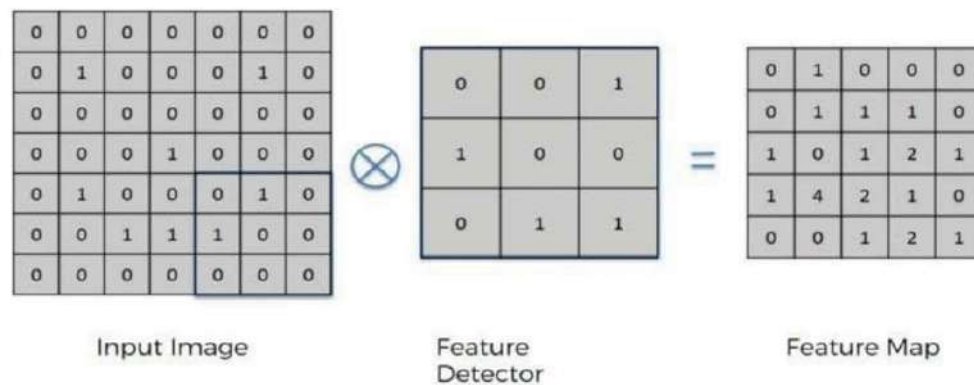
konteks biologis, setiap neuron dalam otak manusia memiliki bidang reseptifnya sendiri dan saling terhubung dengan neuron lainnya sehingga mencakup seluruh lapangan visual. Bidang reseptif merujuk pada bagian kecil dalam lapangan visual di mana setiap neuron dalam sistem penglihatan biologis merespons rangsangan dengan spesifik.

Dalam konteks Convolutional Neural Network (CNN), konsep bidang reseptif ini diadopsi. Setiap neuron dalam CNN menganalisis data hanya dalam wilayah reseptifnya sendiri. Lapisan-lapisan dalam CNN diprogram untuk mengidentifikasi pola yang lebih sederhana, seperti garis dan lengkungan, sebelum mengarah ke pola yang lebih kompleks seperti wajah dan objek. Dengan pendekatan ini, CNN mampu memahami hierarki pola dalam gambar dan meresponsnya secara bertahap. Oleh karena itu, dapat diargumentasikan bahwa penggunaan CNN dapat memberikan kemampuan penglihatan kepada komputer, mirip dengan cara otak manusia memproses informasi visual dengan melibatkan berbagai lapisan neuron. CNN telah membuktikan diri sebagai alat yang efektif dalam pengenalan gambar, termasuk kemampuan untuk mengenali pola dan objek dalam data visual (Bhatt dkk, 2021).

B. Convolution Layer

Convolution Layer adalah lapisan paling dasar tetapi sekaligus paling penting di CNN. Ini pada dasarnya menggulung atau mengalikan matriks pixel yang dihasilkan untuk gambar atau objek yang diberikan untuk menghasilkan *activation map* untuk gambar yang diberikan.

Keuntungan utama dari *activation map* adalah bahwa peta ini menyimpan semua fitur pembeda dari gambar yang diberikan sementara pada saat yang sama mengurangi jumlah data yang akan diproses. Matriks dengan datanya berbelit-belit dan kompleks adalah pendeteksi fitur yang pada dasarnya adalah sekumpulan nilai yang kompatibel dengan mesin. Versi gambar yang berbeda dihasilkan menggunakan nilai pendeteksi fitur yang berbeda. Model berbelit-belit juga dilatih dengan *backpropagation* untuk memastikan kesalahan minimal di setiap lapisan. Menurut set kesalahan terendah, kedalaman dan *padding* diatur (Ajit dkk., 2020).



Gambar 2. 12 Cara kerja Konvolusi

Gambar 2.5 di atas menunjukkan cara kerja konvolusi. Langkah ini melibatkan konvolusi matriks yang berisi data gambar dan kemudian detektor fitur yang memberi kita peta aktivasi atau *feature map*. Apa yang terjadi dalam konvolusi adalah bahwa nilai pada posisi identik dalam data dan peta fitur yaitu nilai yang memiliki nilai 1 atau lebih dari 1 disimpan sementara sisanya dihapus. Matriks dari data citra dibandingkan 3x3 sekaligus. Ukuran fitur detektor

bervariasi dengan jenis CNN yang digunakan. Misalnya ada versi CNN yang menggunakan filter skala 5x5 atau bahkan 7x7 untuk konvolusi. Konvolusi mengikuti

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (2.1)$$

Keterangan:

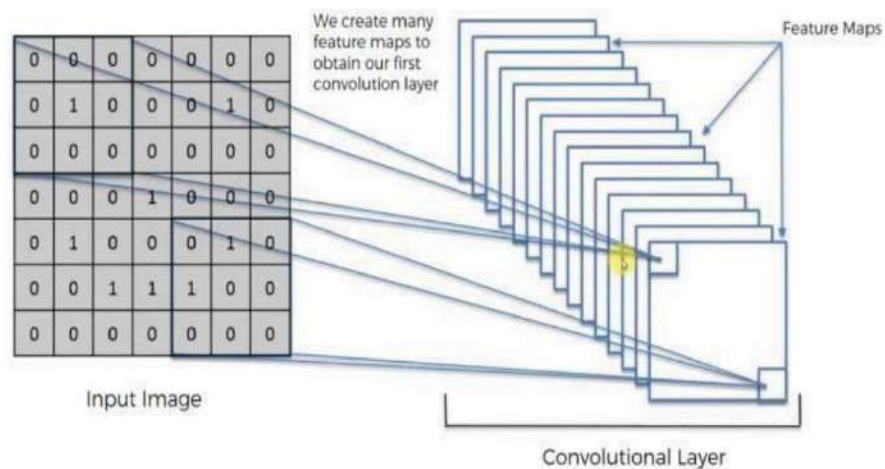
$(f * g)$ = fungsi yang sedang terkonvolusi

t = variable bilangan riil f dan g

$g(\tau)$ = konvolusi fungsi $f(t)$

$d\tau$ = turunan pertama dari fungsi $d(\tau)$

yang bertujuan untuk menunjukkan bagaimana satu fungsi memodifikasi bentuk yang lain. Menjelaskan gambar di atas, data yang dihasilkan untuk gambar ini dimodifikasi menggunakan filter untuk menghasilkan peta aktivasi.



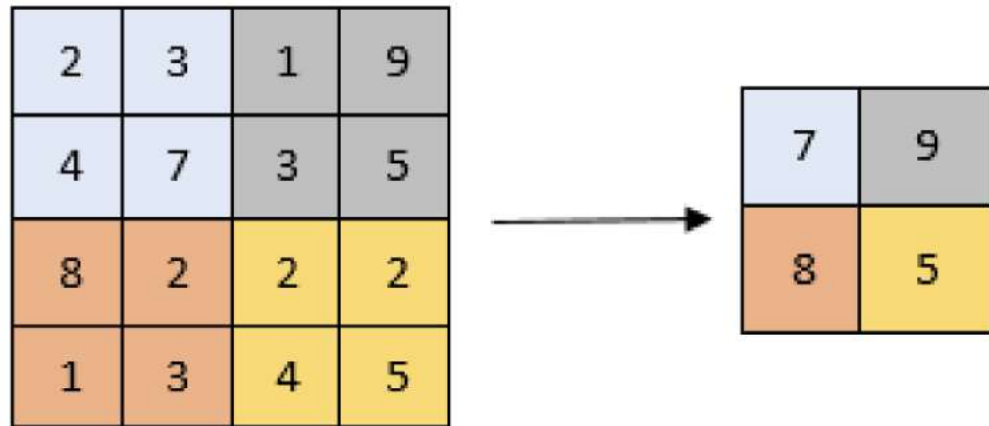
Gambar 2. 13 Layer Konvolusi

C. Pooling Layer

Setelah mendapatkan peta fitur, diperlukan untuk menambahkan lapisan pooling (sub-sampling) dalam CNN, sejajar dengan lapisan konvolusi. Tugas dari lapisan pooling adalah untuk mengurangi ukuran spasial dari fitur hasil konvolusi. Hasil dari pengurangan dimensi ini adalah pengurangan daya komputasi yang diperlukan untuk memproses data. Ini juga membantu dalam ekstraksi karakteristik utama yang invarian terhadap posisi dan rotasi, yang mempertahankan pelatihan model yang efektif. Pooling mengurangi waktu pelatihan sambil mencegah overfitting. Terdapat dua bentuk pooling: pooling maksimum dan pooling rata-rata.

Maximum Pooling

Tensor adalah input ke lapisan pooling. Sebuah kernel berukuran $n \times n$ (sebagai contoh 2×2 dalam contoh yang disebutkan sebelumnya) digerakkan melintasi matriks dalam kasus pooling maksimum, seperti yang diilustrasikan dalam Gambar 2.7, dan nilai maksimum dipilih dan ditempatkan pada lokasi yang sesuai dalam matriks keluaran.

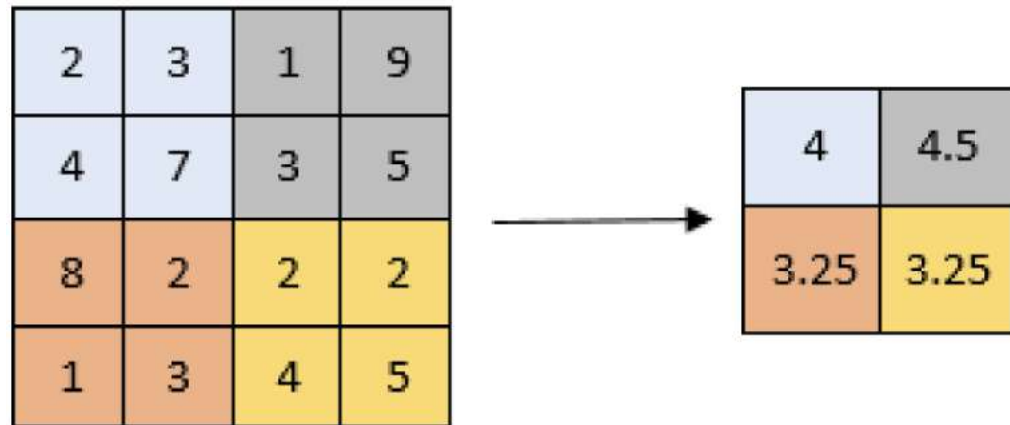


Gambar 2. 14 Max Pooling dengan 2x2 filter dan stride 2

Average Pooling

Average Pooling, sebuah kernel berukuran $n \times n$ digerakkan melintasi matriks, dan rata-rata dari semua nilai dihitung untuk setiap titik dan ditempatkan pada posisi yang sesuai dalam matriks keluaran. Hal ini diulang untuk setiap saluran tensor masukan. Akibatnya, kita memiliki tensor keluaran. Penting untuk diingat bahwa, meskipun pooling mengurangi tinggi dan lebar citra, jumlah saluran (kedalaman) tetap sama. Lapisan pooling menghitung statistik ringkasan dari output sekitarnya untuk menggantikan output jaringan di titik-titik tertentu. Dengan demikian, ini membantu mengurangi dimensi spasial representasi, yang mengurangi jumlah komputasi dan bobot yang diperlukan. Prosedur pooling dilakukan secara independen pada setiap potongan representasi. Rata-rata tetangga persegi, norma L2 dari tetangga persegi, dan rata-rata tertimbang tergantung pada jarak dari piksel tengah adalah semua fungsi pooling seperti yang ditunjukkan dalam Gambar 2.8. Metode yang paling

umum, bagaimanapun, adalah Maximum Pooling, yang melaporkan output terbesar dalam tetangga tersebut (Bhatt dkk, 2021).



Gambar 2. 15 Average Pooling

D. Activation Layer

Activation Layer memainkan peran penting dalam lapisan CNN. Keluaran dari filter diberikan kepada fungsi matematis lain yang disebut fungsi aktivasi. ReLu, yang merupakan singkatan dari rectified linear unit, adalah fungsi aktivasi paling umum yang digunakan dalam ekstraksi fitur CNN. Tujuan utama dari penggunaan fungsi aktivasi adalah untuk menyimpulkan keluaran jaringan saraf, seperti ya atau tidak. Fungsi aktivasi memetakan nilai keluaran antara -1 hingga 1 atau 0 hingga 1, dan sebagainya (tergantung pada fungsi aktivasi yang digunakan) (Mercioni dkk, 2020).

a) Sigmoid Function

Fungsi sigmoid adalah pilihan yang sesuai ketika fungsi output mengambil nilai antara 0 dan 1. Fungsi ini digunakan, misalnya, untuk tugas-tugas klasifikasi dalam jaringan saraf buatan. Fungsi ini digunakan secara intensif dalam dasar Machine Learning, terutama untuk regresi logistik dan beberapa implementasi lainnya. Namun demikian, fungsi ini juga memiliki kelemahan-kelemahan.

Persamaan sigmoid adalah sebagai berikut:

$$f(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

Mulai dari persamaan ini, kita dapat dengan mudah menghitung fungsi turunan sebagai berikut:

$$f' = f(x)(1 - f(x)) \quad (2.3)$$

Seperti yang dapat kita lihat di atas, turunan dari suatu fungsi dapat dihitung dengan mudah, tetapi sifat ini menyebabkan model kehilangan fitur data dan menyebabkan penurunan performa model. Kelemahan utama dari fungsi sigmoid terkait dengan kegunaannya dalam algoritma backpropagation (Mercioni dkk, 2020).

b) ReLu (Rectified Linear Unit)

ReLU, diperkenalkan oleh Krizhevsky dkk. pada tahun 2012, adalah fungsi linear yang mengembalikan input jika inputnya positif, jika tidak, mengembalikan 0. Dalam Stochastic Gradient Descent dengan backpropagation kesalahan untuk melatih model, diperlukan sebuah fungsi yang berperilaku seperti fungsi linear, tetapi merupakan fungsi non-linear yang

memungkinkan pembelajaran hubungan data yang kompleks. Dan ReLU melakukan persis hal ini.

Turunan dari fungsi ReLU mudah dihitung. Turunannya adalah kemiringan. Untuk nilai negatif, kemiringannya sama dengan 0, dan untuk nilai positif, kemiringannya adalah 1 (Mercioni dkk, 2020).

c) *PreLu (Parametric Linear Unit)*

PReLU (Parametric Rectified Linear Unit) adalah sebuah jenis fungsi aktivasi dalam jaringan saraf yang berasal dari fungsi ReLU (Rectified Linear Unit). Salah satu perbedaan utama antara PReLU dan ReLU adalah bahwa PReLU memiliki parameter yang dapat dipelajari (parameter parametrik) untuk mengendalikan kecepatan aktivasi fungsi. Dalam PReLU, selain mengaktifkan unit jika inputnya positif, ada parameter yang memungkinkan penyesuaian tingkat kemiringan pada bagian negatif input.

Penggunaan PReLU memungkinkan jaringan saraf untuk lebih fleksibel dalam pembelajaran dan memungkinkan adaptasi yang lebih baik terhadap data pelatihan. Hal ini juga membantu mengurangi masalah yang sering terjadi dengan ReLU, yaitu "dying ReLU," di mana neuron berhenti aktif selama pelatihan karena kemiringan negatifnya tetap 0. PReLU membantu mengatasi masalah ini dengan memungkinkan kemiringan yang dapat diatur untuk nilai-nilai negatif.

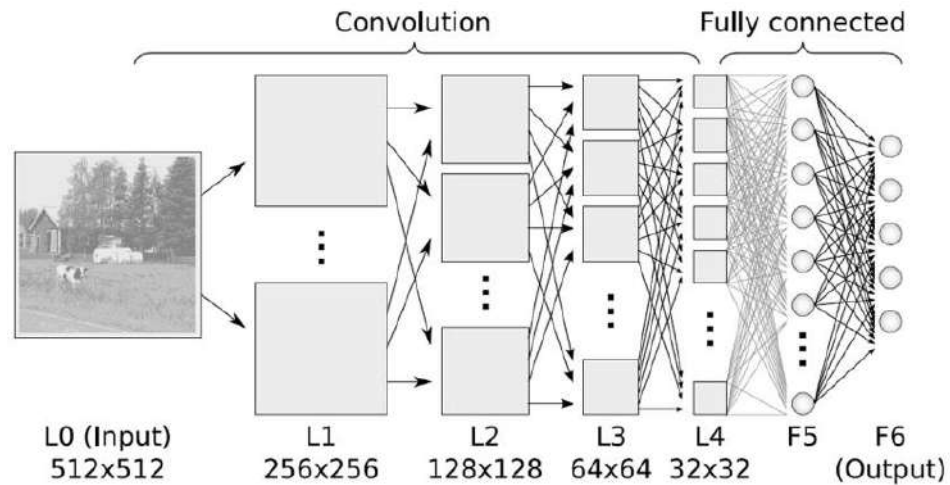
Dengan kata lain, PReLU memperkenalkan parameter-parameter tambahan yang memungkinkan model untuk memutuskan sendiri kemiringan

yang optimal untuk setiap neuron, sehingga meningkatkan kinerja dan stabilitas selama pelatihan.

E. Fully Connected Layer

Misalnya *Fully Connected Layer* mirip dengan cara neuron diatur dalam jaringan saraf tradisional. Oleh karena itu, setiap *node* di lapisan yang terhubung penuh terhubung langsung ke setiap *node* di lapisan sebelumnya dan di lapisan berikutnya seperti yang ditunjukkan pada Gambar 2.9. Dari gambar ini kita dapat mencatat bahwa masing-masing *node* di *frame* terakhir di *pooling* lapisan terhubung sebagai vektor ke lapisan pertama dari lapisan yang sepenuhnya terhubung. Ini adalah parameter yang paling banyak digunakan dengan CNN dalam lapisan ini, dan membutuhkan waktu lama dalam proses pelatihan.

Kelemahan utama dari *Fully Connected Layer*, adalah bahwa ia mencakup banyak parameter yang memerlukan komputasi kompleks dalam contoh *training*. Oleh karena itu, perlu untuk menghilangkan jumlah *node* dan koneksi. *Node* dan koneksi yang dihapus dapat dipenuhi dengan menggunakan teknik *dropout*. Misalnya, LeNet dan AlexNet merancang jaringan yang dalam dan luas sambil menjaga kompleks komputasi tetap konstan



Gambar 2. 16 Contoh Fully Connected Layer

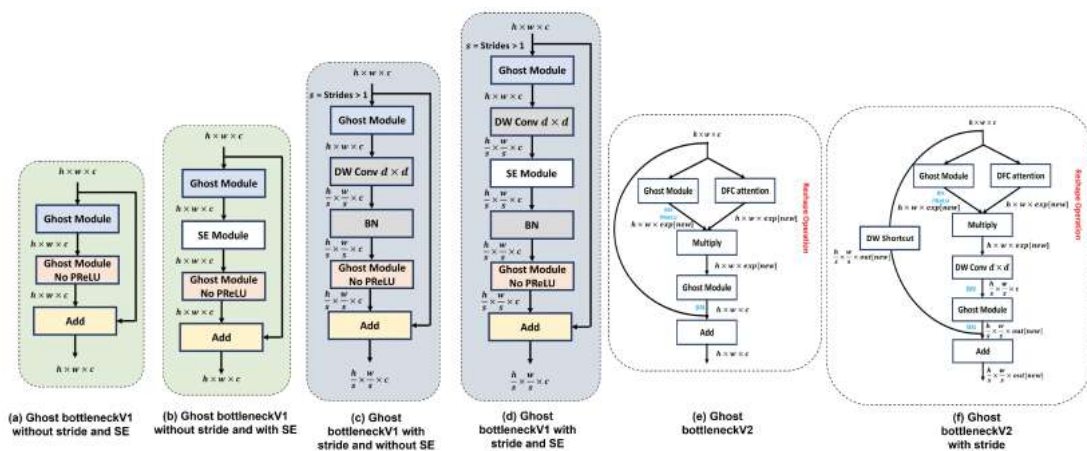
2.1.15. Algoritma GhostFaceNets

GhostFaceNets, yang dikembangkan oleh Qian et al. pada tahun 2020 (Qian et al, 2020), merupakan sebuah arsitektur jaringan saraf konvolusi yang dirancang khusus untuk tugas pengenalan wajah. GhostFaceNets mencapai performa yang tinggi dalam deteksi wajah dan ekstraksi fitur wajah dengan menggunakan pendekatan holistik dan toleran terhadap degradasi citra. Arsitektur GhostFaceNets terdiri dari beberapa lapisan konvolusi dan penggabungan yang digunakan untuk menghasilkan fitur wajah yang diskrit. Arsitektur GhostFaceNets memberikan keunggulan dalam memproses citra wajah dengan presisi tinggi dan toleransi terhadap variasi kondisi lingkungan, seperti pencahayaan yang rendah atau perubahan sudut pandang.

Dalam penelitian yang dilakukan oleh Hu et al. pada tahun 2021 [2], GhostFaceNets digunakan untuk deteksi wajah pada video CCTV dengan akurasi yang tinggi. Hasil penelitian tersebut menunjukkan bahwa

GhostFaceNets mampu mengenali wajah dengan tepat bahkan dalam situasi yang penuh dengan gangguan seperti perubahan pencahayaan dan pergerakan objek.

Selain itu, penelitian yang dilakukan oleh Zhang et al. pada tahun 2022 (Zhang et al, 2022) menunjukkan bahwa GhostFaceNets juga dapat digunakan untuk mengenali individu dalam dataset yang besar dengan performa yang tinggi. Penelitian tersebut membuktikan bahwa GhostFaceNets memiliki kemampuan pengenalan yang akurat dan efisien dalam mengidentifikasi individu berdasarkan fitur-fitur wajah. Gambar 2.4 menunjukkan beberapa arsitektur GhostFaceNet.



Gambar 2. 17 Beberapa arsitektur GhostFaceNets bekerja

2.1.16. Algoritma Deepsort

DeepSort (*Deep association metric of Simple Online and Real time Tracking*) adalah pendekatan berbasis *Deep Learning* untuk melacak kustom

objek dalam video (Wojke et al., 2017). Dalam penelitian awal mengenai DeepSort, Deepsort digunakan untuk melacak individu yang ada dalam rekaman video pengawasan. Metode ini bekerja dengan memanfaatkan pola yang dipelajari melalui objek yang terdeteksi di gambar yang kemudian digabungkan dengan informasi temporal untuk memprediksi lintasan terkait objek yang menarik. Ini akan melacak setiap objek yang dipertimbangkan dengan pemetaan pengenalan unik untuk analisis statistik lebih lanjut. Deepsort juga berguna untuk menangani tantangan terkait seperti oklusi, berbagai sudut pandang, kamera non-stasioner, dan anotasi data pelatihan.

Untuk pelacakan yang efektif, Kalman-Filter (R. E. KALMAN, n.d.) dan algoritma Hungaria digunakan. Kalman-Filter secara rekursif digunakan untuk asosiasi yang lebih baik, dan dapat memprediksi posisi masa depan berdasarkan posisi saat ini. Sedangkan Algoritma Hungaria digunakan untuk asosiasi dan atribusi id yang mengidentifikasi apakah suatu objek di frame saat ini sama dengan frame sebelumnya. Dalam mendeteksi, deepsort menggambarkan setiap target dengan ruang delapan dimensi dalam model kecepatan konstan linier, sebagai berikut: dimana (u, v) mewakili lokasi piksel horizontal dan vertikal dari pusat target, λ dan h masing-masing mewakili skala (Real) dan rasio aspek kotak pembatas target. Perlu diingat bahwa rasio aspek dianggap konstan. Variabel lainnya adalah kecepatan masing-masing variabel. Ketika deteksi dikaitkan dengan target, *bounding box* yang terdeteksi digunakan untuk memperbarui status target dimana komponen kecepatan diselesaikan secara optimal melalui kerangka Kalman-Filter. Nantinya, standar Kalman-Filter

digunakan dengan kecepatan gerak konstan dan linier model observasi, dimana koordinat pembatas (u , v , λ , h) diambil sebagai pengamatan langsung dari keadaan objek. Jika tidak ada deteksi yang dikaitkan dengan target, keadaannya diprediksi tanpa koreksi menggunakan model kecepatan linier (persamaan 2.1).

$$x = [u, v, \lambda, h, x', y', \lambda', h']^T \quad (2.1)$$

u, v = posisi *centroid* setiap *bounding box* objek
 λ = nilai aspek rasio *bounding box* objek
 h = tinggi *bounding box*
 x', y', λ', h' = *velocity factors* setiap variable-variabel
 sebelumnya dalam koordinat gambar (berkaitan dengan

Skenario tracking didefinisikan dalam ruang delapan dimensi (u , v , λ , h , x' , y' , λ' , h'). Dimana nilai u , v menyatakan posisi *centroid* dari setiap *bounding box*, nilai λ merepresentasikan nilai rasio tinggi-lebar *bounding box*, dan h merepresentasikan tinggi *bounding box*. Untuk setiap track T , mulai dari asosiasi pengukuran terakhir yang berhasil λ_T , jumlah total frame dihitung. Dengan prediksi positif dari Kalman-Filter, penghitung bertambah dan kemudian ketika trek dikaitkan dengan pengukuran, Nilainya diatur ulang ke 0. Selanjutnya, jika trek yang diidentifikasi melebihi waktu maksimum yang telah ditentukan, maka objek tersebut dianggap telah meninggalkan adegan dan trek yang sesuai akan dihapus dari set trek. Dan jika tidak ada trek yang tersedia untuk beberapa objek yang terdeteksi maka hipotesis trek baru dimulai untuk

setiap trek tak dikenal dari objek baru yang terdeteksi yang tidak dapat dipetakan ke trek yang ada. Untuk tiga frame pertama, trek baru diklasifikasikan sebagai tidak terbatas sampai pemetaan pengukuran yang berhasil dihitung. Jika trek tidak berhasil dipetakan dengan pengukuran maka itu akan dihapus dari kumpulan trek. Algoritma Hungaria (Kuhn, 1955) kemudian digunakan untuk memecahkan masalah pemetaan antara pengukuran yang baru tiba dan keadaan Kalman yang diprediksi dengan mempertimbangkan informasi gerakan dan tampilan dengan bantuan *Mahalanobis-Distance* yang dihitung di antara objek yang dapat di formulakan menjadi (persamaan 2.2):

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \quad (2.2)$$

$d^{(1)}(i, j)$	= nilai sampel Mahalanobis-Distance
d_j	= nilai bounding box terdistribusi
y_i	= nilai rata-rata sample pada tiga frame pertama
S_i	= kovarians sampel (dengan bias terkoreksi)

dimana proyeksi distribusi lintasan ke-i ke dalam ruang pengukuran diwakili oleh (y_i , S_i) dan deteksi kotak pembatas ke-j oleh d_j . Jarak Mahalanobis mempertimbangkan ketidakpastian ini dengan memperkirakan jumlah standar deviasi, deteksi jauh dari lokasi trek rata-rata. Selanjutnya, dengan menggunakan matrik ini, dimungkinkan untuk mengecualikan asosiasi yang tidak mungkin dengan membatasi jarak Mahalanobis. Keputusan ini dilambangkan dengan indikator yang bernilai 1 jika hubungan antara jalur ke-i dan deteksi ke-j dapat diterima dengan persamaan 2.3:

$$b_{i,j}^{(1)} = 1[d^{(1)}(i,j) < t^{(1)}] \quad (2.3)$$

$b^{(1)}(i,j)$ = nilai normalisasi sampel Mahalanobis-Distance

$d^{(1)}(i,j)$ = nilai sampel Mahalanobis-Distance

t^l = Mahalanobis threshold (9.4877)

meskipun jarak Mahalanobis bekerja dengan efisien tetapi gagal di lingkungan dimana gerakan kamera terjadi, oleh karena itu matrik lain diperkenalkan untuk masalah penugasan. Ini matrik kedua mengukur jarak kosinus terkecil antara trek ke- i dan deteksi ke- j di ruang tampilan sebagai berikut:

$$d^{(2)}(i,j) = \min\{1 - r_j^T r_k^{(i)} \mid r_k^{(i)} \in \mathbb{R}^2\} \quad (2.4)$$

r_j = konstanta descriptor penampil pada deteksi ke- j

r_k = konstanta descriptor penampil pada tracking ke- j

\mathbb{R}^2 = Keseluruhan nilai matrik jarak kosinus terkecil

dan ambang batas yang sesuai diukur untuk indikator ini pada kumpulan data pelatihan terpisah. Untuk membangun masalah asosiasi, kedua matrik digabungkan menggunakan jumlah berbobot:

$$c_{i,j} = \lambda d^{(1)}(i,j) + (1 - \lambda) d^{(2)}(i,j) \quad (2.5)$$

$c_{i,j}$ = jumlah bobot gabungan matriks jarak kosinus terkecil dan matriks jarak Mahalanobis

dimana asosiasi dapat diterima jika berada di dalam gerbang wilayah kedua matrik:

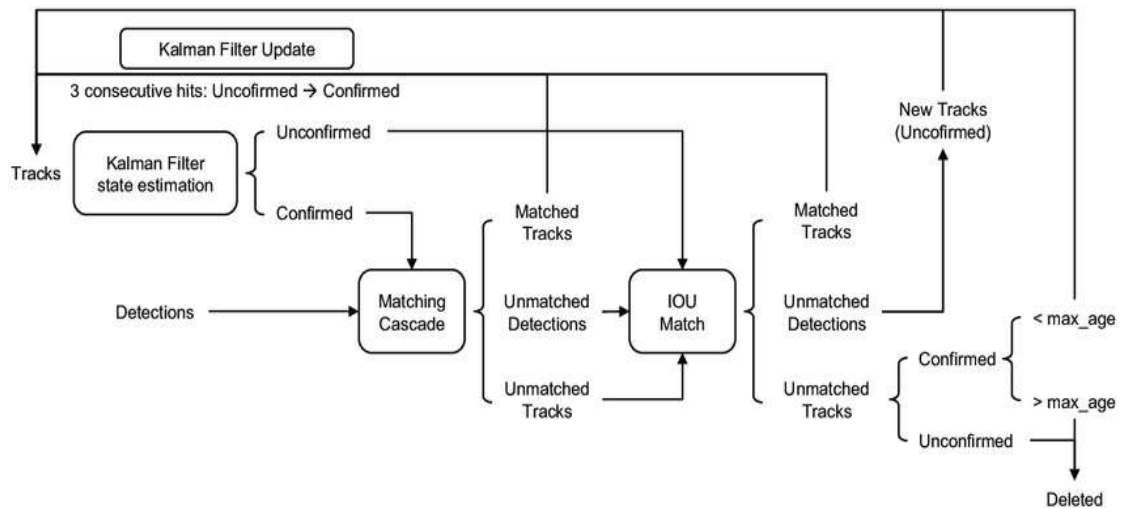
$$b_{i,j} = \prod_{m=1} 2b_{i,j}^{(m)}. \quad (2.6)$$

$b_{i,j}$ = matriks gerbang wilayah jarak kosinus terkecil dan matriks jarak Mahalanobis
 \prod = produk cartesian

Pengaruh setiap matrik pada biaya asosiasi gabungan dapat dikontrol melalui hyperparameter λ diberi kode warna berdasarkan asosiasinya dengan index setiap objek. Dalam menetapkan deteksi ke target yang ada, masing-masing target geometri kotak pembatas diperkirakan dengan memprediksi lokasi dalam bingkai saat ini. Matriks biaya penugasan adalah kemudian dihitung sebagai jarak persimpangan-over-union atau *Intersection-of-Union* (IOU) antara setiap deteksi dan semua kotak pembatas yang diprediksi dari target yang ada. Tugas diselesaikan secara optimal menggunakan algoritma Hungaria. Selain itu, minimal IOU dikenakan untuk menolak tugas dimana deteksi untuk target tumpang tindih kurang dari IOUmin.

Jarak IOU dari kotak pembatas secara implisit menangani oklusi jangka pendek yang disebabkan oleh target yang lewat. Khususnya, ketika target ditutupi oleh penghalang objek, hanya occluder yang terdeteksi, karena jarak IOU tepat mendukung deteksi dengan skala yang sama. Hal ini memungkinkan kedua target occluder dikoreksi dengan deteksi sementara target yang dicakup tidak terpengaruh karena tidak ada penetapan terbuat. Gambar 2.5 berikut

menggambar secara sederhana bagaimana deepSort bekerja (Avola et al., 2021).



Gambar 2. 18 Alur kerja DeepSORT

2.2. Penelitian Terkait

Penelitian sebelumnya yang terkait dengan prediksi daerah rawan kecelakaan ialah sebagai berikut:

2. Deep Learning based Loitering Detection System using Multi-camera Video Surveillance Network (Rashmiranjan Nayak dkk., 2019)

Rashmiranjan Nayak pada tahun 2019 menggunakan YoloV3 dan DeepSORT yang dilengkapi dengan kemampuan identifikasi ulang (ReID) untuk membuat sistem yang dapat mendeteksi dan melacak pola pergerakan gelandangan pada sebuah wilayah tertentu menggunakan dua CCTV yang

saling terintegrasi. Sistem yang diusulkan memberikan akurasi yang ditingkatkan sebesar 96 % pada rata-rata fps 33 (tanpa ReID) dan 81,5 % pada rata-rata fps 30 (dengan ReID).

3. A survey on deep learning based face recognition (Guodong Guo. dkk, 2019)

Guodong Guo pada tahun 2019 melakukan review penggunaan deep learning untuk studi kasus face recognition mendapatkan beberapa kesimpulan diantaranya, penggunaan dan pengembangan algoritma deep learning untuk menyelesaikan face recognition didominasi oleh CNN dan beberapa varian turunannya, hal ini dikarenakan CNN berhasil meringkas data gambar menjadi representasi lebih sederhana.

4. AN EFFECTIVE IMPLEMENTATION OF FACE RECOGNITION USING DEEP CONVOLUTIONAL NETWORK (Hayder Najm dkk, 2019)

Hayder Najm dkk menerapkan menerapkan CNN dengan Fuzzy logic untuk melakukan rekognisi wajah. Hasil dari inovasi ini terbukti menaikkan akurasi CNN pada dataset yang awalnya 78.39 menjadi 98.12 dengan diikuti penurunan execution time yang awalnya 9.12 secon menjadi 5.83 secon.

5. Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility (Chen Z. dkk, 2019)

Zhihao Chen dkk membandingkan SSD dan YOLO V3 dalam mendeteksi, tracking dan estimasi jarak kendaraan, mendapatkan hasil bahwa YOLO V3 lebih baik dalam akurasi dan dapat beradaptasi pada kondisi real-time.

7. Person Re-Identification using Deep Learning Networks: A Systematic Review (Ankit Yadav dkk., 2019)

Ankit Yadav dkk dalam papernya yang berjudul Person Re-Identification using Deep Learning Network: A Systematic Review mencoba membandingkan kinerja beberapa algoritma Deep Learning dalam kasus pengenalan individu yang sama yang diambil dari beberapa lokasi atau waktu yang berbeda. Salah satu hasil yang didapatkan bahwa pendekatan algoritma GAN (Generative Adversarial Network) sangat membantu untuk menyelesaikan kebutuhan style transfer pada pendekatan Re-Identification.

2.3. State Of The Art

Tabel 2. 1 State of The Art

No	Nama Peneliti	Judul & Tahun	Sumber Data	Metode & Algoritma	Hasil
1	Edwin Jose, Greeshma M., Mithun Haridas (2019)	Face Recognition based Surveillance System Using FaceNet and MTCNN on Jetson TX2 (Jose et al., 2019)	data latih yang digunakan berupa foto, sementara untuk data uji menggunakan data foto dan vide	FaceNet dan MTCNN pada Jetson TX2	Deteksi akurasi mencapai 97%
2	Rashmiranjan Nayak, Mohini Mohan Behera, V Girish, Umesh Chandra Pati and Santos Kumar Das (2019)	Deep Learning based Loitering Detection System using Multi-camera Video Surveillance Network (Nayak et al., 2019)	PETS 2007 dataset	YoloV3 & DeepSORT	Sistem yang diusulkan memberikan peningkatan akurasi sebesar 96% pada rata-rata fps 33 (tanpa ReID) dan 81,5% pada rata-rata fps 30 (dengan ReID)
.3	Sapna Sharma, Vikrant Dhama (2020)	Abnormal Human Behavior Detection in Video Using	UMN Dataset	Metode deteksi perbedaan temporal pada frame	Algoritma yang diusulkan membutuhkan komputasi yang lebih sedikit dengan menggunakan

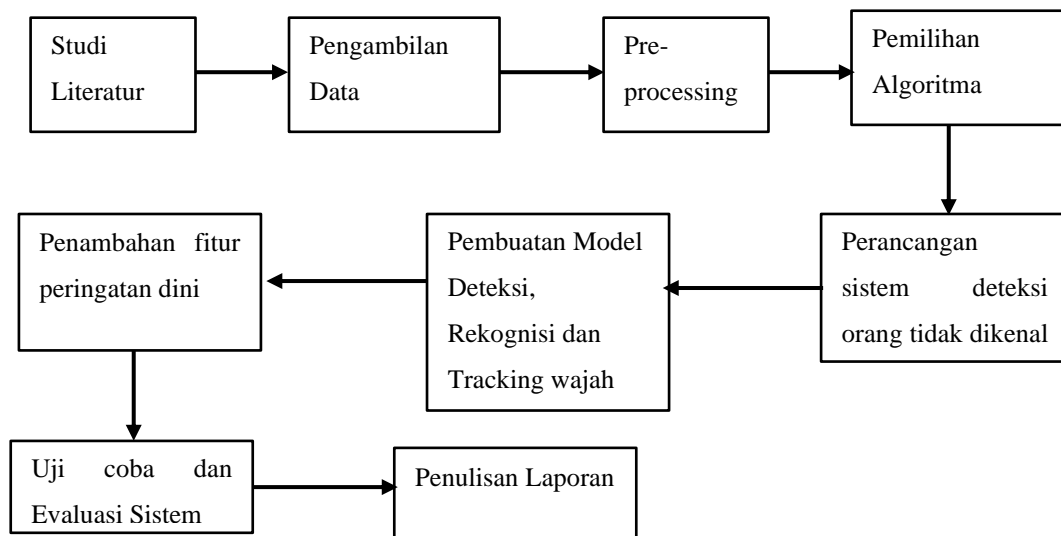
		Suspicious Object Detection (Sharma et al., 2020)			pendekatan yang efisien dan juga meningkatkan akurasi dalam video sistem pengawasan untuk mendeteksi aktivitas abnormal di lingkungan real-time
4	Nicolai Wojke, Alex Bewley, Dietrich Paulus (2017)	SIMPLE ONLINE AND REALTIME TRACKING WITH A DEEP ASSOCIATION METRIC (2017	MOT 16 Dataset, MARS Dataset	DeepSORT	Evaluasi eksperimental menunjukkan bahwa metode yang diusulkan mengurangi jumlah penggantian id pada objek yang sama sebesar 45%, mencapai kinerja kompetitif secara keseluruhan pada frekuensi gambar tinggi
5	MOHAMAD ALANSARI, OUSSAMA ABDUL HAY, SAJID JAVED, ABDULHADI SHOUFAN, YAHYA ZWEIRI, AND NAOUFEL WERGH (2023)	GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations (Alansari et al., 2023)	MS1MV2 dataset, MS1MV3 dataset.	GhostFace Nets	Evaluasi GhostFaceNets menggunakan berbagai tolok ukur menunjukkan bahwa model ini menawarkan kinerja yang unggul sementara membutuhkan kompleksitas komputasi sekitar 60-275 MFLOP. Ini secara signifikan lebih rendah dari beberapa algoritma face detection yang sudah ada. Akurasi tertinggi dari algoritma ini mencapai 99.65

BAB III

METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini akan berfokus pada proses pre-processing dan augmentasi pada data tiap gambar wajah yang akan dimasukkan di database. Hal ini bertujuan untuk mengoptimalkan proses pendeteksian meskipun terjadi perubahan keadaan seperti perubahan intensitas cahaya maupun perubahan cuaca. Selain itu penelitian ini juga mencoba menerapkan multi-face recognition dengan oklusi masker medis dan kacamata hitam, hal ini berkaitan dengan kondisi individu di sekolah yang akan menjadi tempat mengambil data. Setelah model face detection telah mencapai kondisi-kondisi yang telah dijabarkan, selanjutnya akan dilakukan penambahan fitur notifikasi dengan beberapa ketentuan. Gambar 3.1 menunjukkan bagaimana alur penelitian berjalan:



Gambar 3. 1 Alur Penelitian

3.2. Waktu dan Lokasi Penelitian

Penelitian ini dilakukan selama 3 bulan dimulai pada 5 Juni 2023 hingga September 2023. Penelitian dilakukan dengan mengambil data wajah siswa, guru di salah satu kelas di SD Telkom Makassar. Adapun pelaksanaan penelitian dilakukan di Lab AIMP, Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Perubahan lokasi pengambilan data yang awalnya direncanakan di SMP 6 Makassar sebagaimana yang tercantum pada proposal awal, berubah menjadi SD Telkom Makassar mempertimbangkan ketersediaan infrastruktur seperti jaringan internet yang secara real-time dapat dihubungkan di CCTV. SMP 6 Makassar, belum memiliki fasilitas internet yang dapat kami gunakan sedangkan SD Telkom Makassar dapat meminjamkan akses infrastruktur tersebut

3.3. Populasi dan Sampel

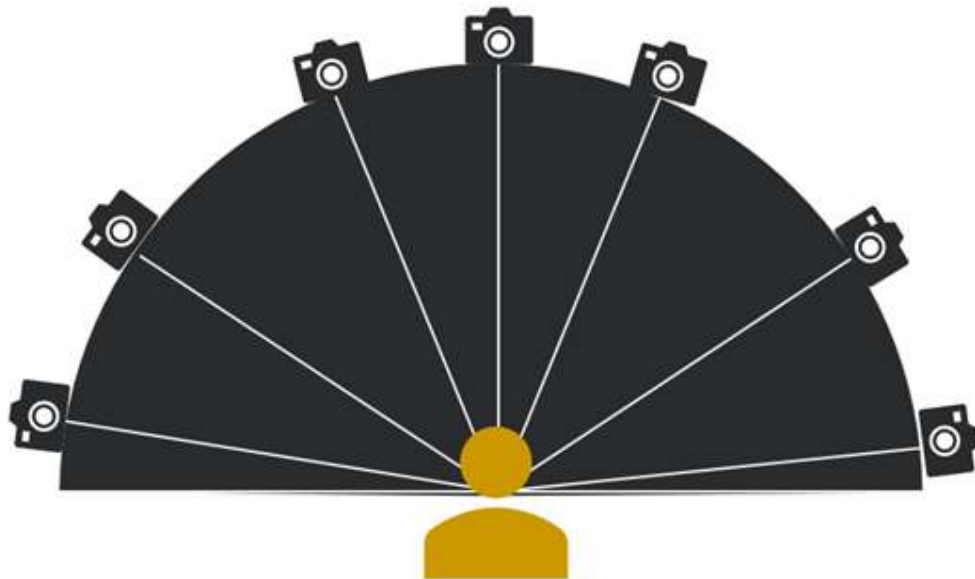
Populasi dari penelitian ini adalah sebagian guru, siswa yang saling berasosiasi satu sama lain. Ini mencakup setiap individu pada setiap tingkat di SD Telkom Makassar. Sampel dari penelitian ini adalah beberapa guru, siswa dan penjemput siswa yang saling berasosiasi.

Pemilihan sampel akan mengikuti arahan pada pihak-pihak berwenang dilingkungan SD Telkom Makassar dengan jumlah total individu yang akan dijadikan sampel adalah 10 – 20 individu. Hal ini dikarenakan. Hal ini dikarenakan purwarupa pada penelitian ini tidak membutuhkan metode pengambilan data tertentu.

Pada pelaksanaan lapangan kami berhasil mengambil 10 data siswa dan 8 data wajah guru yang kemudian kami asosiasikan dalam proses pengkodean.

3.4. Teknik Pengambilan data

Data yang digunakan pada penelitian ini merupakan data primer yang langsung diambil di lokasi penelitian. Data akan diambil dengan menggunakan kamera dengan 7 sudut pandang terhap objek; 5, 30, 65, 90, 115, 150, 175 derajat kemiringan terhadap objek. Secara sederhana, proses pengambilan data dapat digambarkan dengan gambar 3.2 dan 3.3 berikut:

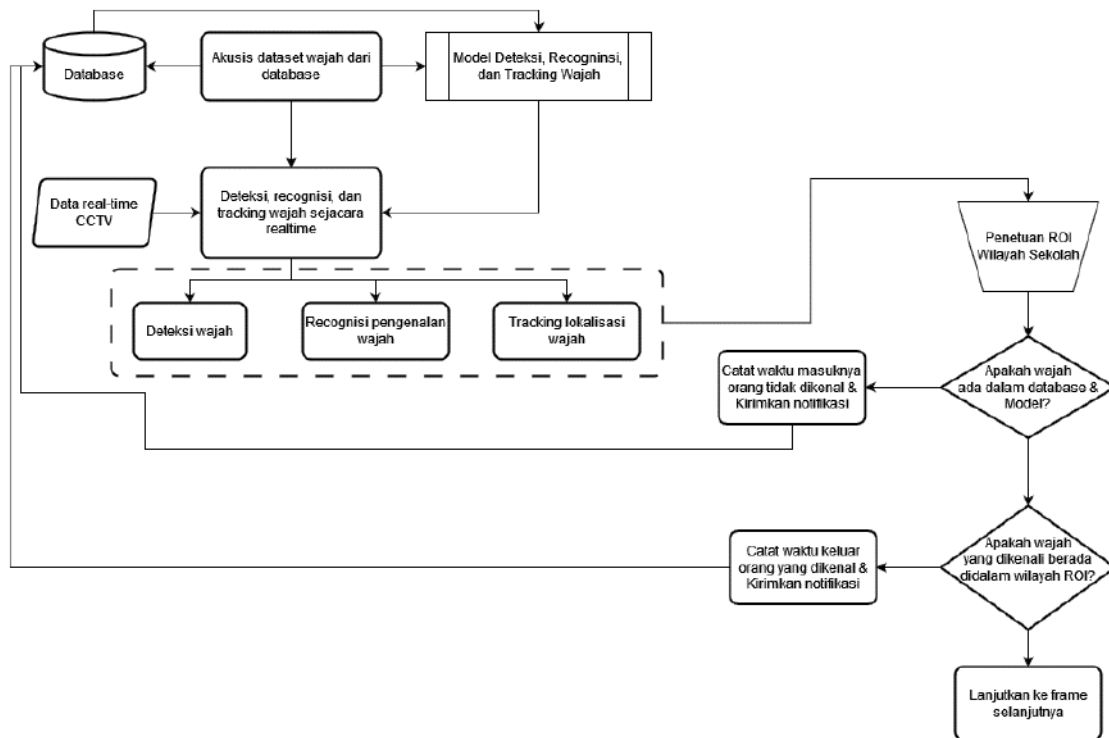


Gambar 3. 2 Gambaran Teknik Pengambilan data



Gambar 3. 3 Contoh ilustrasi pengaturan kamera

3.5. Desain Usulan sistem



Gambar 3. 4 Desain Umum Sistem

Penjelasan system:

Sistem deteksi, pengenalan, dan tracking wajah yang diimplementasikan menggunakan model GhostFaceNets dan DeepSORT bertujuan untuk mengenali dan melacak wajah manusia dalam data streaming yang diperoleh melalui CCTV. Sistem ini juga mampu melakukan klasifikasi wajah dengan membandingkan data wajah yang terdeteksi dengan database yang berisi nama-nama yang terkait. Berikut adalah langkah-langkah umum yang dilakukan oleh sistem:

1. Inisialisasi ROI (Region of Interest):

- o ROI adalah wilayah sekolah yang ditentukan sebelumnya dalam citra atau video dari CCTV. ROI dapat meliputi lapangan dan gerbang sekolah

- Sistem melakukan inisialisasi ROI dengan menggunakan teknik pemrosesan citra seperti segmentasi atau anotasi manual untuk memisahkan objek manusia dan wajah dari latar belakang.

2. Deteksi Wajah:

- Sistem menggunakan model GhostFaceNets untuk mendeteksi wajah dalam data streaming yang diterima dari CCTV.
- GhostFaceNets menggunakan jaringan saraf konvolusi untuk mengekstraksi fitur-fitur wajah dan mendeteksi keberadaan wajah dalam citra atau video.
- Setiap wajah yang terdeteksi kemudian diisolasi dan diproses lebih lanjut untuk pengenalan dan tracking.

3. Recognisi Pengenalan Wajah:

- Setelah wajah terdeteksi, sistem menggunakan GhostFaceNets untuk melakukan pengenalan wajah dengan membandingkannya dengan data wajah yang ada dalam database.
- Data wajah dalam database berisi informasi identitas seperti nama-nama yang terkait.
- Jika wajah yang terdeteksi cocok dengan data wajah dalam database, sistem mengklasifikasikan wajah tersebut dengan nama yang sesuai.
- Hasil akhir dari proses ini kemudian akan terbentuk bounding box penanda berwarna hijau jika individu yang ada dalam database dan bounding box berwarna merah jika individu tidak ada dalam database

4. Tracking Wajah menggunakan DeepSORT:

- Setelah pengenalan wajah dilakukan, sistem menerapkan algoritma DeepSORT untuk melacak wajah selama data streaming berlangsung.
- DeepSORT menggunakan pendekatan deteksi dan pelacakan objek yang menggabungkan informasi dari deteksi wajah dan data pergerakan objek dari frame ke frame.

- Algoritma ini memungkinkan sistem untuk mengikuti pergerakan wajah yang terdeteksi dan melacaknya seiring waktu.
- Proses tracking dimaksudkan agar proses deteksi dan rekognisi tidak perlu mengulang dari fase awal pada setiap pergantian frame.

5. Catatan Waktu Masuk dan Keluar:

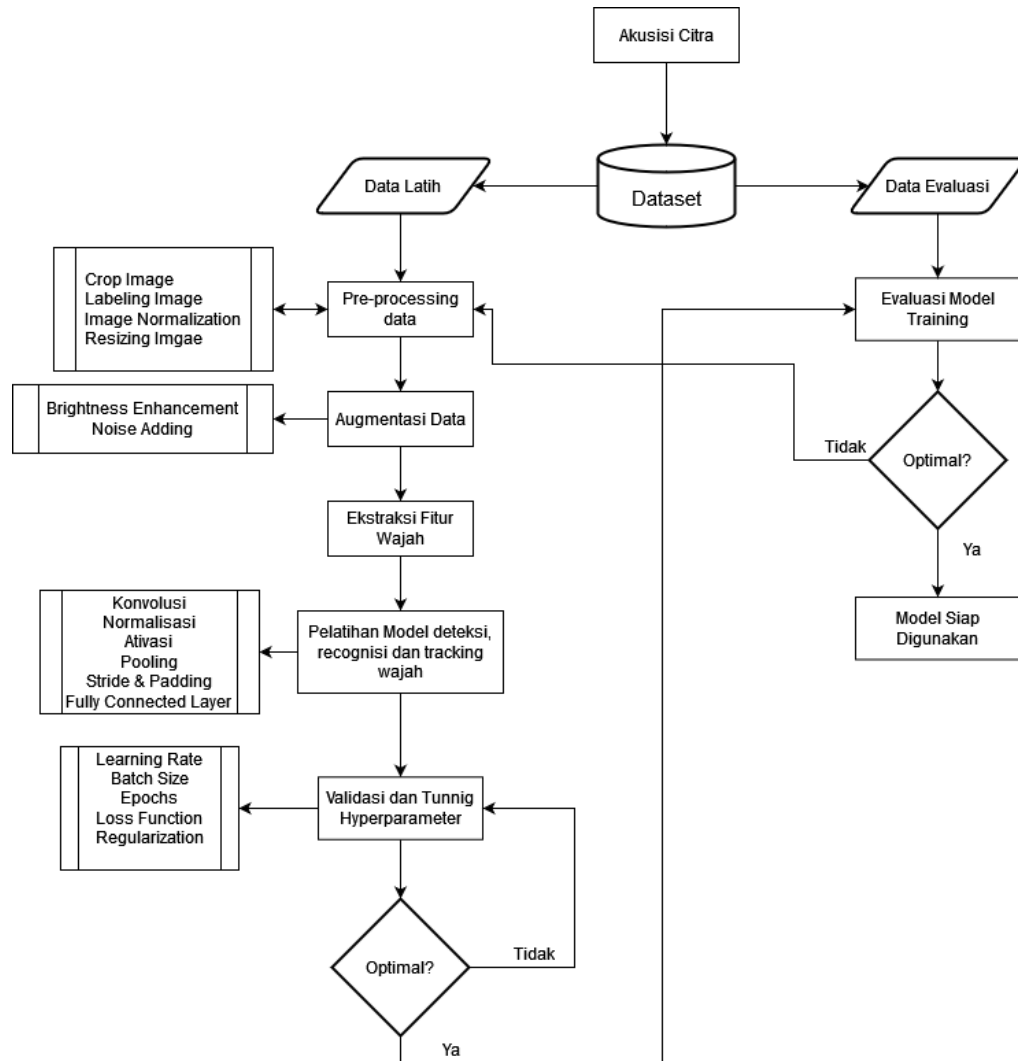
- Sistem mencatat waktu masuk dan keluar setiap wajah baik yang dikenali maupun tidak dikenali berdasarkan data waktu dari video streaming dan dimasukkan kedalam database.
- Informasi waktu masuk dan keluar ini digunakan untuk memantau kapan setiap individu meninggalkan area sekolah.

6. Notifikasi keamanan via WhatsApp (WA):

- Sistem membandingkan data wajah yang dikenali dengan database untuk menentukan apakah individu yang dikenali merupakan orang yang terdaftar atau tidak.
- Jika individu yang terdeteksi tidak dikenal dan tercatat dalam database memasuki wilayah sekolah (ROI), sistem akan mengirimkan notifikasi melalui WhatsApp untuk memberi tahu tentang adanya individu tidak dikenal yang masuk ke wilayah sekolah.
- Selain itu, jika individu yang terdeteksi dan tercatat dalam database meninggalkan ROI wilayah, sistem juga akan mengirimkan notifikasi melalui WhatsApp untuk memberi tahu tentang kepergian individu tersebut.

Sistem ini memadukan kemampuan deteksi wajah GhostFaceNets untuk mengidentifikasi wajah, pelacakan wajah DeepSORT untuk melacak pergerakan individu, dan pengenalan wajah dengan database untuk membandingkan data individu dengan database yang ada. Hal ini memungkinkan sistem untuk melakukan pemantauan kehadiran dan memberikan notifikasi jika ada individu tidak dikenal yang masuk atau keluar dari wilayah sekolah yang ditentukan (ROI).

Proses Pelatihan Model deteksi, rekognisi dan tracking wajah



Gambar 3. 5 Proses Pembuatan Model

Penjelasan pelatihan model:

Setelah data terkumpul, selanjutnya data akan disimpan dalam database dan dibagi menjadi data training dan data evaluasi dengan rasio tertentu. Pada data yang termasuk kategori data latih, akan dilakukan beberapa pre-processing seperti pelabelan dan crop, hal ini bertujuan untuk meringankan proses komputasi. Selain itu,

akan dilakukan proses augmentasi yang bertujuan untuk mensimulasikan kondisi-kondisi kota makassar pada beberapa kejadian, seperti perubahan intensitas cahaya maupun jika terjadi hujan.

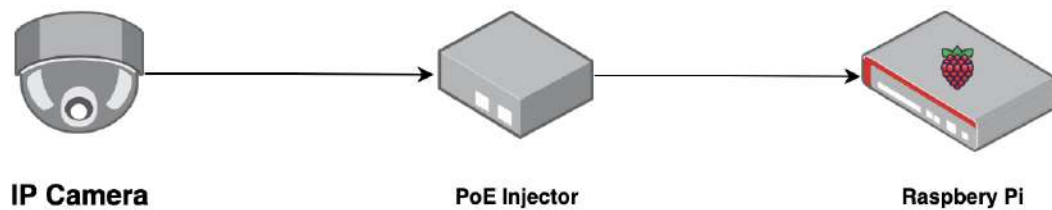
Setelah data latih dianggap siap untuk masuk ke algoritma, selanjutnya dilakukan pelatihan model menggunakan GhostFaceNets untuk penyelesaian kasus deteksi dan recognisi wajah serta algoritma DeepSORT untuk kasus tracking wajah. Pada proses pelatihan ini akan dilakukan beberapa kali proses validasi dan tuning hyperparameter yang bertujuan untuk menemukan bentuk model paling optimal untuk kasus pendeteksian wajah diwilayah makassar.

Setelah model menunjukkan hasil yang optimal, selanjutnya model Kembali akan diuji menggunakan data evaluasi, hasil dari proses ini adalah model yang siap digunakan dan diintegrasikan dengan kamera cctv yang terpasang.

3.6. Perangkat Pengiriman Video

Penggunaan IP kamera sebagai alat utama dalam pemantauan dan pengumpulan data visual dalam sistem ini. IP kamera diintegrasikan ke dalam sistem dengan bantuan PoE Injector, yang berfungsi sebagai pengirim daya dan data melalui satu kabel Ethernet. Dengan demikian, proses instalasi menjadi lebih sederhana karena tidak memerlukan akan kabel daya yang terpisah. Selanjutnya, Raspberry Pi digunakan sebagai perangkat pemrosesan yang memiliki peran penting dalam pengiriman data yang diterima dari IP kamera. Raspberry Pi memberikan platform yang kuat untuk pengiriman, pengambilan keputusan, dan kendali terhadap IP kamera.

Kombinasi antara IP kamera, PoE Injector, dan Raspberry Pi memungkinkan penyediaan solusi yang efisien dan terintegrasi dalam mendeteksi orang tidak dikenal di sekolah. Tampilan rancangan system untuk pengiriman video dapat kita lihat pada Gambar 3.6



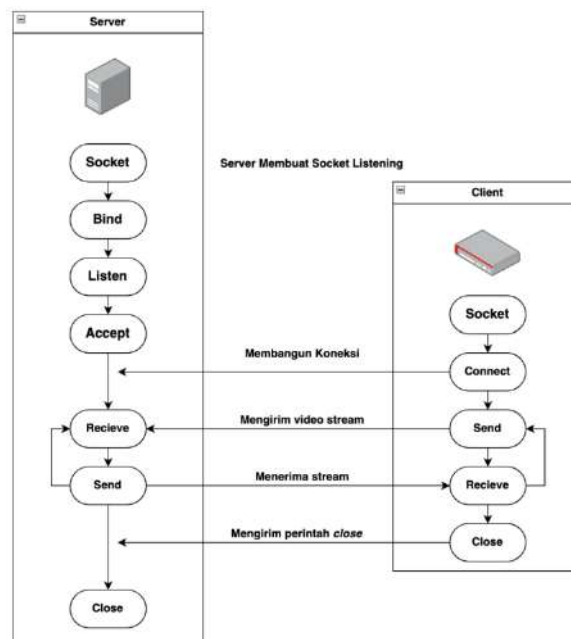
Gambar 3. 6 Desain perangkat pengiriman video streaming

Mengirim video dari Raspberry Pi ke server dengan menggunakan koneksi socket adalah salah satu aplikasi yang mendalam dan penting dalam pengembangan berbagai sistem berbasis komputasi mini. Dalam konteks ini, Raspberry Pi bertindak sebagai perangkat penghasil video yang mengumpulkan data visual melalui kamera terpasangnya. Data video tersebut kemudian diproses dan dipecah menjadi paket-paket yang dapat ditransmisikan melalui koneksi socket. Socket digunakan sebagai saluran komunikasi yang memungkinkan transfer data yang kontinu dan real-time antara Raspberry Pi dan server yang menerima video tersebut.

Koneksi socket ini memanfaatkan protokol komunikasi yang dapat disesuaikan, seperti TCP atau UDP, tergantung pada kebutuhan aplikasi. Protokol TCP menawarkan koneksi yang andal dengan jaminan pengiriman paket data dalam urutan yang benar. Ini cocok untuk situasi di mana keandalan transmisi lebih penting daripada latensi. Di sisi lain, protokol UDP menyediakan transmisi data yang lebih cepat tanpa jaminan urutan paket. Ini lebih sesuai dalam kasus-kasus di mana sedikit

kerugian data dapat diterima demi respon yang lebih cepat. Selain itu, dalam mengirim video, seringkali ada kebutuhan untuk kompresi data untuk mengurangi penggunaan bandwidth dan memastikan transfer yang lebih lancar. Raspberry Pi dapat menggunakan algoritma kompresi video yang sesuai sebelum mengirimkan data video melalui koneksi socket. Kemudian, di sisi server, video tersebut diterima, di-dekompresi, dan dianalisis sesuai kebutuhan aplikasi, seperti pemantauan keamanan atau analisis visual.

Dengan mengimplementasikan koneksi socket yang baik antara Raspberry Pi dan server, kita dapat mencapai transmisi video yang efisien, real-time, dan andal. Hal ini sangat penting dalam berbagai konteks, mulai dari sistem pengawasan keamanan hingga aplikasi berbasis IoT yang melibatkan pengambilan data visual dari Raspberry Pi yang memprosesnya dan mengirimkannya ke server pusat untuk analisis lebih lanjut.



Gambar 3. 7 Cara Kerja Socket

3.7. Teknik Analisa Kinerja Sistem

Analisis kinerja sistem merupakan tahap penting dalam penelitian ini untuk mengevaluasi keefektifan dan kehandalan dari optimalisasi algoritma computer vision yang digunakan, yaitu GhostFaceNets untuk deteksi dan rekognisi wajah serta DeepSORT untuk pelacakan wajah. Analisis kinerja sistem ini bertujuan untuk mengukur performa sistem dalam mendeteksi dan melacak orang tidak dikenal berdasarkan video CCTV dengan kondisi lingkungan yang berbeda di Kota Makassar.

Pertama, untuk menganalisis kinerja GhostFaceNets sebagai algoritma deteksi dan rekognisi wajah, akan dilakukan evaluasi terhadap tingkat akurasi deteksi dan kecepatan pemrosesan dalam mengidentifikasi wajah individu dalam video. Metrik evaluasi yang akan digunakan meliputi tingkat kebenaran deteksi (true positive rate), tingkat kesalahan deteksi (false positive rate), presisi (precision), dan recall. Selain itu metode evaluasi untuk deteksi dan recognisi juga akan menggunakan metode perhitungan accuracy dan ArcFaceLoss.

Selanjutnya, untuk menganalisis kinerja DeepSORT sebagai algoritma pelacakan wajah, akan dilakukan evaluasi terhadap keakuratan pelacakan, kestabilan pelacakan, dan kecepatan pelacakan. Metrik evaluasi yang akan digunakan meliputi akurasi pelacakan (tracking accuracy), kehilangan pelacakan (tracking loss), dan kecepatan pemrosesan.

Analisis kinerja sistem akan dilakukan dengan menggunakan video CCTV secara realtime yang menggambarkan situasi lingkungan Kota Makassar yang berbeda, termasuk kondisi pencahayaan yang berubah-ubah, kepadatan orang yang bervariasi, dan gerakan yang kompleks. Pada setiap kondisi, sistem akan dijalankan

dan hasil deteksi, rekognisi, dan pelacakan akan dievaluasi menggunakan metrik evaluasi yang telah disebutkan sebelumnya.

Hasil analisis kinerja sistem akan memberikan informasi penting mengenai keefektifan dan kehandalan dari algoritma yang digunakan. Dengan memperoleh pemahaman yang mendalam tentang performa sistem, penelitian ini dapat mengidentifikasi kelebihan dan kekurangan algoritma yang digunakan, serta mengevaluasi sejauh mana algoritma tersebut dapat digunakan dalam mendeteksi dan melacak orang tidak dikenal dalam lingkungan CCTV di Kota Makassar.

Dalam penelitian ini, analisis kinerja sistem akan menjadi landasan yang kuat untuk mengukur keberhasilan optimalisasi algoritma computer vision dalam konteks deteksi, rekognisi, dan pelacakan wajah. Hasil analisis ini akan membantu peneliti dalam menyimpulkan apakah algoritma GhostFaceNets dan DeepSORT dapat diandalkan untuk mendeteksi orang tidak dikenal berbasis video CCTV dalam kondisi lingkungan yang beragam di wilayah sekolah di Kota Makassar.

3.8. Instrumen Penelitian

Adapun beberapa instrument penelitian yang digunakan pada penelitian ini adalah:

1. Software
 - Windows Home 11x64
 - Google Collaboratory
 - Python 3.9 (64 Bit)
 - Python 3.9.12 64-bit

- TensorFlow 2.8.0 (CUDA compatible GPU needed for GPU training)
- Keras 2.8.0
- keras_cv_attention_models
- glob2
- pandas
- tqdm
- sklearn
- scikit-image
- cuda_11.6.r11.6/compiler.31057947_0
- pywhatkit
- socket

2. Hardware

- Laptop Acer Nitro AN515-N1
- Processor Intel Core i9-11900H
- CCTV HIKVISION
- GPU NVIDIA GeForce RTX 3060
- RaspberryyPi 4 8GB
- IP Camera 2 MP

BAB IV

HASIL DAN PEMBAHASAN

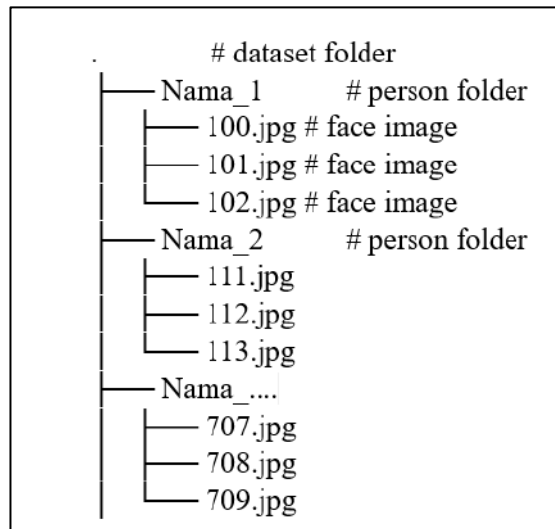
4.1. Tahap Pengambilan Data

Pengambilan data dilakukan di SD Telkom Makassar pada 27 Juli 2023, Data yang digunakan pada penelitian ini merupakan data primer berupa 10 wajah siswa dan 5 wajah orangtua/guru yang langsung diambil di lokasi penelitian. Data akan diambil dengan menggunakan kamera dengan 7 sudut pandang terhadap objek; 5, 30, 65, 90, 115, 150, 175 derajat kemiringan terhadap objek.



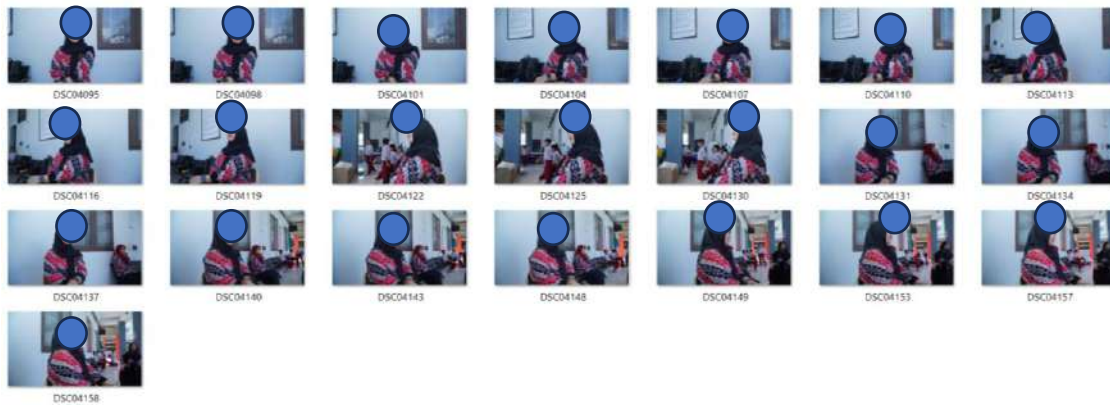
Gambar 4. 1 Teknik Pengambilan Data

Hasil dari data tersebut dikelompokkan dalam folder-folder berdasarkan nama pemilik wajah. Struktur data dapat digambarkan pada gambar 4.2 berikut:



Gambar 4. 2 Struktur Penyimpanan data

Berikut merupakan gambar contoh hasil pengambilan gambar dalam 1 folder:

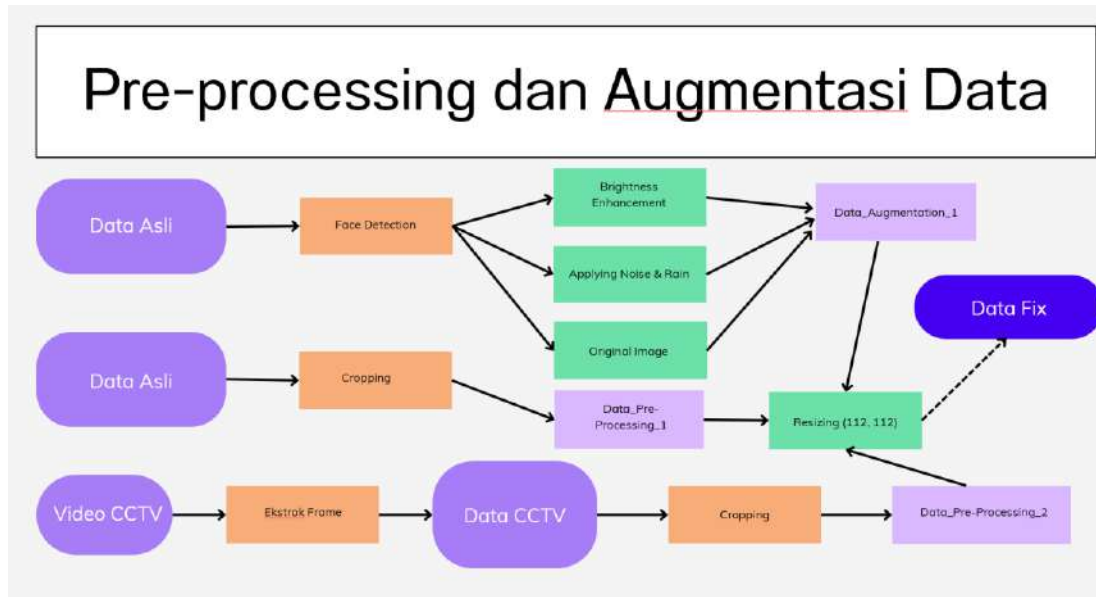


Gambar 4. 3 Contoh Data Asli Hasil Pengambilan Data

Data raw yang telah diambil kemudian memasuki tahap pre-processing dan augmentasi.

4.2. Pre-processing dan Augmentasi Data

Pada tahap pre-processing dan augmentasi, data yang ada disiapkan dan divariasikan agar siap diolah oleh algoritma. Pada penelitian ini, tahap pre-processing meliputi crop dan resizing, augmentasi meliputi penambahan tingkat keterangan dan kegelapan pada gambar, serta proses penambahan filter hujan. Gambar 4.4 berikut menggambarkan proses pre-processing dan augmentasi yang dilakukan pada gambar:kegelapan pada gambar, serta proses penambahan filter hujan. Gambar berikut menggambarkan proses pre-processing dan augmentasi yang dilakukan pada gambar:



Gambar 4. 4 Keseluruhan proses pre-processing dan augmentasi pada data

Pada gambar 4.4 tersebut beberapa rangkaian pre-processing dan augmentasi data dilakukan terpisah dan memanfaatkan data mentah atau data raw. Brightness Enhancement bertujuan untuk mengatur tingkat gelap terang pada gambar, hal ini untuk menangani perubahan intensitas cahaya pada kamera CCTV bergantung pada kondisi ruangan dan waktu terjadinya proses deteksi dan rekognisi wajah. Penerapan filter hujan dan penambahan beberapa noise dimaksudkan agar proses pendeteksian dan recognisi tetap berjalan optimal meskipun terjadi perubahan cuaca seperti hujan. Proses pre-processing dan augmentasi dilakukan secara manual maupun menggunakan proses pengkodean dengan bantuan library *opencv_python*, hasil dari proses ini kemudian digabungkan kedalam file data raw berdasarkan nama dari pemilik data asli. Hasil dari pre-processing dapat dilihat pada gambar 4.5 berikut:



Gambar 4. 5 Data hasil pre-processing dan augmentasi

Setelah data dikumpulkan dalam satu folder yang sama, kemudian dilakukan proses penyamaan ukuran pixel menjadi 112 x 112, hal ini agar algoritma dapat Belajar dengan proporsi ukuran yang sama dan seragam. Data raw dan data hasil pre-processing dan augmentasi kemudian akan dipisahkan, hal ini agar data raw dapat dikonfigurasi ulang bilamana proses augmentasi dan pre-processing belum menunjukkan hasil yang optimal.

4.3. Tahap Pembangunan Model GhostFaceNet

Setelah dari 18 wajah sudah dilakukan proses pre-processing dan augmentasi, selanjutnya masuk ke tahap Pembangunan model menggunakan algoritma GhostFaceNets. Algoritma ini sendiri memiliki 2 versi arsitektur, versi 1 dan versi 2. Perbedaan utama dari kedua arsitektur ini adalah GhostFaceNet menggunakan penambahan cabang DFC attention pada saat proses pelatihan data, hal ini dikarenakan cabang DFC attention sangat efisien dan mampu menangkap ketergantungan jarak jauh antara piksel yang terletak di lokasi spasial yang berbeda.

Kedua arsitektur ini kemudian dilakukan proses pelatihan menggunakan data disertai penyesuaian parameter dan hyperparameter. Dalam proses membangun model, beberapa hal yang perlu diperhatikan adalah:

1. Pemilihan tipe model dan arsitektur

Pemilihan tipe dan model arsitektur berkaitan dengan keadaan data serta kemampuan infrastruktur yang dimiliki. Pada penelitian ini, GhostFaceNets versi 1 menunjukkan hasil yang lebih baik dan optimal dalam memprediksi data. Hal ini disebabkan penggunaan cabang DFC attention ternyata menyebabkan penghematan komputasi yang berakibat pada kurang ter-representasinya data yang hendak dilatih dikarenakan ketergantungan antara pixel dihilangkan, sedangkan gambar yang digunakan sudah cukup mencakup spesifik daerah wajah.

2. Jumlah Dimensi embedding_layer

Embedding layer merupakan representasi gambar kedalam bentuk matriks, semakin besar dimensi embedding layer yang ditetapkan maka semakin banyak informasi dari gambar yang ditampung, namun ini dapat menyebabkan banyaknya informasi kurang berguna dari gambar serta membutuhkan kekuatan komputasi yang lebih banyak. Setelah dilakukan beberapa percobaan, didapatkan bahwa jumlah dimensi embedding layer yang terbaik adalah 512 untuk tiap arsitektur.

3. Penambahan regularisasi L2.

Regularisasi L2 adalah atau *Ridge Regression* adalah metode untuk menambahkan penalti saat kompleksitas model meningkat. Parameter

regularisasi (λ) menghukum semua parameter kecuali intersep sehingga model menggeneralisasi data dan tidak akan overfit. Overfit adalah keadaan dimana model akan melakukan generalisasi pada tiap objek, sebagai contoh jika terdapat 18 objek saat proses pelatihan, keadaan overfit menyebabkan model akan condong untuk memprediksi salah satu objek secara berlebihan. Hal ini disebabkan kompleksitas layer pada arsitektur.

Penambahan regularisasi L2 dilakukan dengan syntax berikut:

```
if output_weight_decay >= 1:
    l2_weight_decay = 0
    for ii in self.basic_model.layers:
        #Look for "regularizer" in model's layers
        if hasattr(ii, "kernel_regularizer") and
isinstance(ii.kernel_regularizer, keras.regularizers.L2):
            l2_weight_decay = ii.kernel_regularizer.l2
            break
        print(">>>> L2 regularizer value from basic_model:",
l2_weight_decay)
    output_weight_decay *= l2_weight_decay * 2
    self.output_weight_decay = output_weight_decay
```

4. Penambahan Fungsi Aktivasi

Funsgi aktivasi merupakan fungsi yang digunakan pada jaringan saraf untuk mengaktifkan atau tidak mengaktifkan neuron. Karakteristik yang harus dimiliki oleh fungsi aktivasi jaringan perambatan balik antara lain harus kontinyu, terdiferensialkan, dan tidak menurun secara monotonis.

Fungsi aktivasi diterapkan pada output setiap neuron dalam lapisan jaringan saraf dan memungkinkan sinyal untuk mengalir melalui jaringan dengan cara yang kompleks dan non-linear. Tujuan utama dari fungsi aktivasi

adalah untuk memperkenalkan jaringan saraf untuk menangkap pola dan fitur yang lebih kompleks dalam data, yang tidak mungkin dicapai dengan transformasi linier saja.

Pada penelitian ini fungsi aktivasi yang digunakan adalah PReLU. Fungsi aktivasi Parametric Rectified Linear Unit (PReLU) adalah variasi dari fungsi aktivasi ReLU (Rectified Linear Unit) yang memiliki parameter tambahan yang memungkinkan gradien yang lebih halus pada nilai negatif. PReLU diperkenalkan untuk mengatasi masalah "dying ReLU" yang dapat terjadi saat pelatihan jaringan saraf tiruan menggunakan fungsi ReLU.

Berikut adalah cara kerja PReLU:

1. Definisi Fungsi Aktivasi PReLU:

Fungsi aktivasi PReLU didefinisikan sebagai berikut

$$f(x) = x, \text{ jika } x \geq 0$$

$$f(x) = a * x, \text{ jika } x < 0$$

Di sini, a adalah parameter yang dapat diatur selama pelatihan (berupa parameter tambahan dalam PReLU dibandingkan dengan ReLU). Pada penelitian ini parameter a berganti-ganti dan bervariasi mengikuti parameter lain seperti learning rate, fungsi regularisasi L2, dan parameter-parameter lainnya

2. Kelebihan Parameter a :

Perbedaan utama antara PReLU dan ReLU adalah penggunaan parameter a . Nilai a merupakan skalar positif yang bisa dipelajari oleh model selama pelatihan. Saat $a = 0$, PReLU berperilaku seperti ReLU biasa. Namun,

jika a diberi nilai positif, PReLU memungkinkan gradien non-nol pada nilai negatif, mengatasi masalah "dying ReLU".

3. Keuntungan PReLU:

- Mengatasi Masalah Dying ReLU:

Fungsi PReLU membantu menghindari masalah "dying ReLU", di mana neuron dengan nilai negatif yang besar tidak mengalami perubahan gradien dan tidak belajar selama pelatihan. Dengan PReLU, gradien dapat mengalir melalui bagian negatif dari fungsi, memungkinkan pembelajaran yang lebih baik.

- Regularisasi Implisit:

Parameter a yang dapat diatur memberikan pengaturan tambahan yang dapat membantu dalam regularisasi model. Model dapat memutuskan apakah menggunakan a yang lebih kecil atau mendekati 0 untuk menghindari nilai negatif yang signifikan.

4. Pelatihan dan Pembelajaran Nilai a :

Selama pelatihan, parameter a akan diatur sedemikian rupa sehingga model dapat memutuskan apakah harus menggunakan gradien positif atau gradien yang lebih halus pada nilai negatif. Nilai a akan diubah secara iteratif untuk mengoptimalkan kinerja model pada data pelatihan.

Dalam konteks model GhostFaceNets yang ada pada penelitian ini, lapisan embedding berperan dalam menghasilkan representasi fitur yang lebih baik, regularisasi L2 membantu mengendalikan kompleksitas model, dan penggantian

ReLU dengan PReLU menciptakan non-linearitas yang lebih kompleks dalam model. Semua ini merupakan teknik untuk mengoptimalkan kinerja dan kemampuan umum model Anda dalam memodelkan data.

4.4. Tahap Pengujian Model GhostFaceNet

Dalam proses pelatihan model, setiap model yang dihasilkan di tiap iterasi akan dievaluasi sebanyak tiga kali. Evaluasi pertama adalah bagaimana algoritma tersebut menetapkan proses evaluasi bagi hasil pembelajarannya sendiri. Pada evaluasi tahap pertama ini, nilai akurasi dan loss akan menjadi salah satu parameter bagi algoritma dalam proses pembelajaran, dari proses tersebut kemudian dilakukan penyesuaian parameter dan hyperparameter seperti *learning rate*. Learning rate merupakan tingkat kepekaan algoritma dalam mempelajari data, semakin kecil nilai learning rate, maka proses pelatihan akan semakin mendetail namun hal ini dapat menyebabkan overfitting disertai kebutuhan komputasi yang besar.

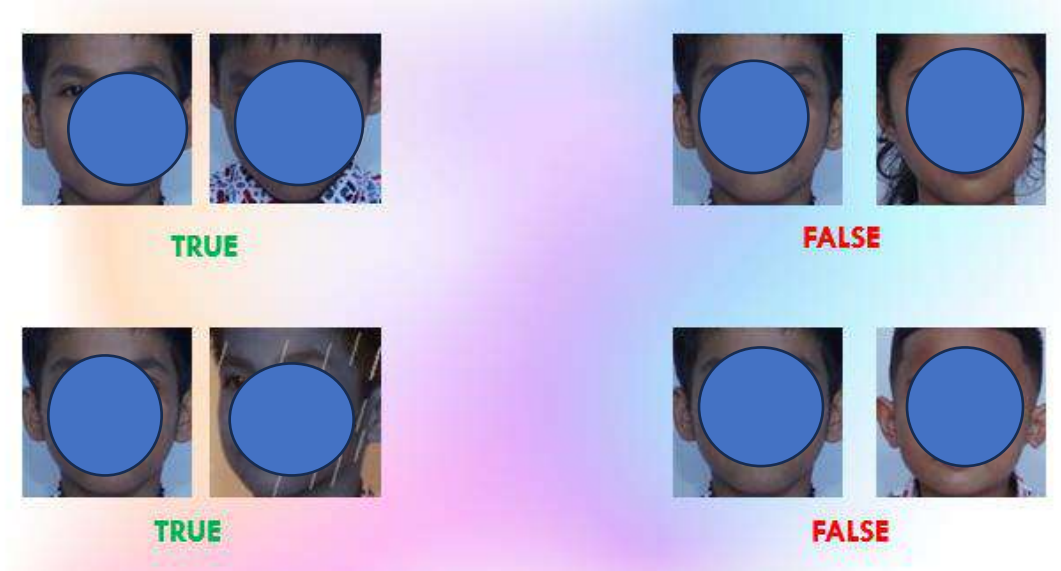
Selain tahap algoritma menguji hasil penelitiannya sendiri, penelitian ini juga membuat file berformat .bin yang menyimpan informasi mengenai kumpulan kombinasi gambar yang bernilai benar dan salah. Proses ini melibatkan pengkodean dengan bantuan tensorflow berikut:

```
# Save to bin
import pickle
import tensorflow as tf
from skimage.io import imread

image_list = ...
is_same_list = ...

bb = [tf.image.encode_jpeg(imread(ii)).numpy() for ii in
image_list]
with open("data_fix.bin", "wb") as ff:
    pickle.dump([bb, is_same_list], ff)
```

Pada kode tersebut, *image_list* berisi list nama gambar yang dipasangkan dan ditentukan secara acak, kemudian setiap satu pasang gambar pada tiap urutan akan bernilai *True* atau *False* yang disimpan pada list *is_same_list*. File *.bin* dapat diilustrasikan gambar:

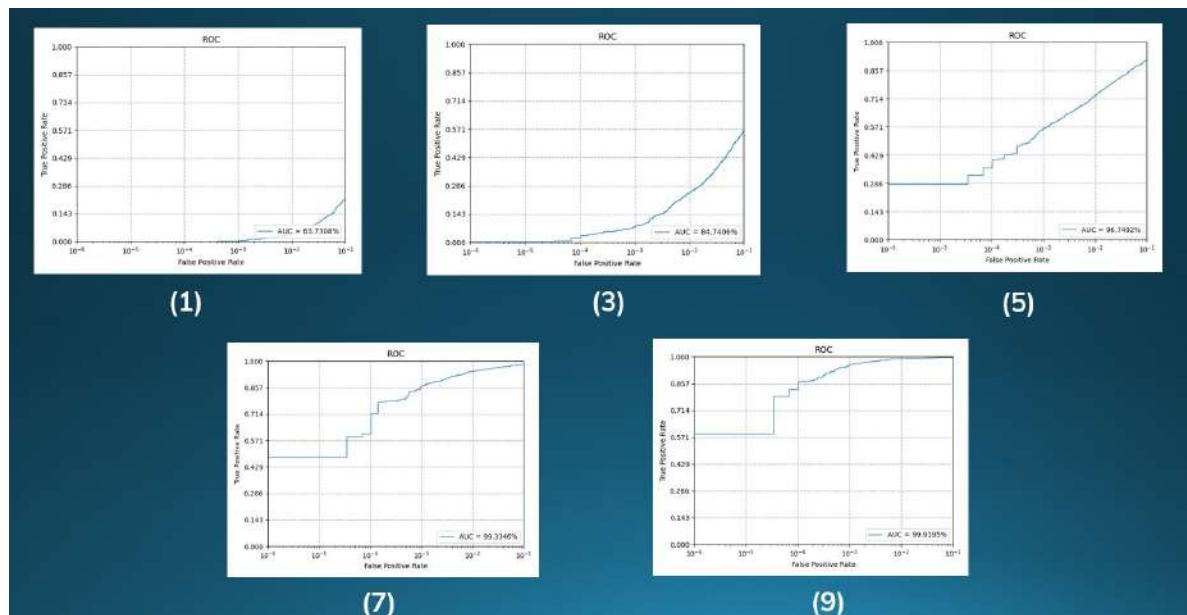


Gambar 4. 6 Data Format *.bin*

Proses pelatihan kemudian akan melibatkan tunning atau mencari parameter dan hyhperparameter terbaik bagi proses pembelajaran dan model hasil

pembelajaran. Setiap proses training dilakukan iterasi sebanyak 200 kali, dengan satu kali proses training membutuhkan waktu sekitar satu hingga lima jam, bergantung pada struktur model, parameter dan hyperparameter yang diuji. Semakin kompleks struktur model, maka proses training menggunakan *Google Collaboratory* akan semakin lama, begitupun lama tidaknya proses pelatihan bergantung pada besar kecilnya setiap parameter dan hyperparameter.

Setelah dilakukan proses training, setiap iterasi kemudian diamati grafik hasil pembelajarannya. Gambar 4.7 berikut merupakan beberapa contoh grafik akurasi hasil pembelajaran algoritma pada salah satu sesi pelatihan:



Gambar 4. 7 Contoh Grafik akurasi hasil pembelajaran pada tiap iterasi

4.5. Tahap Pengujian Berdasarkan Data Video

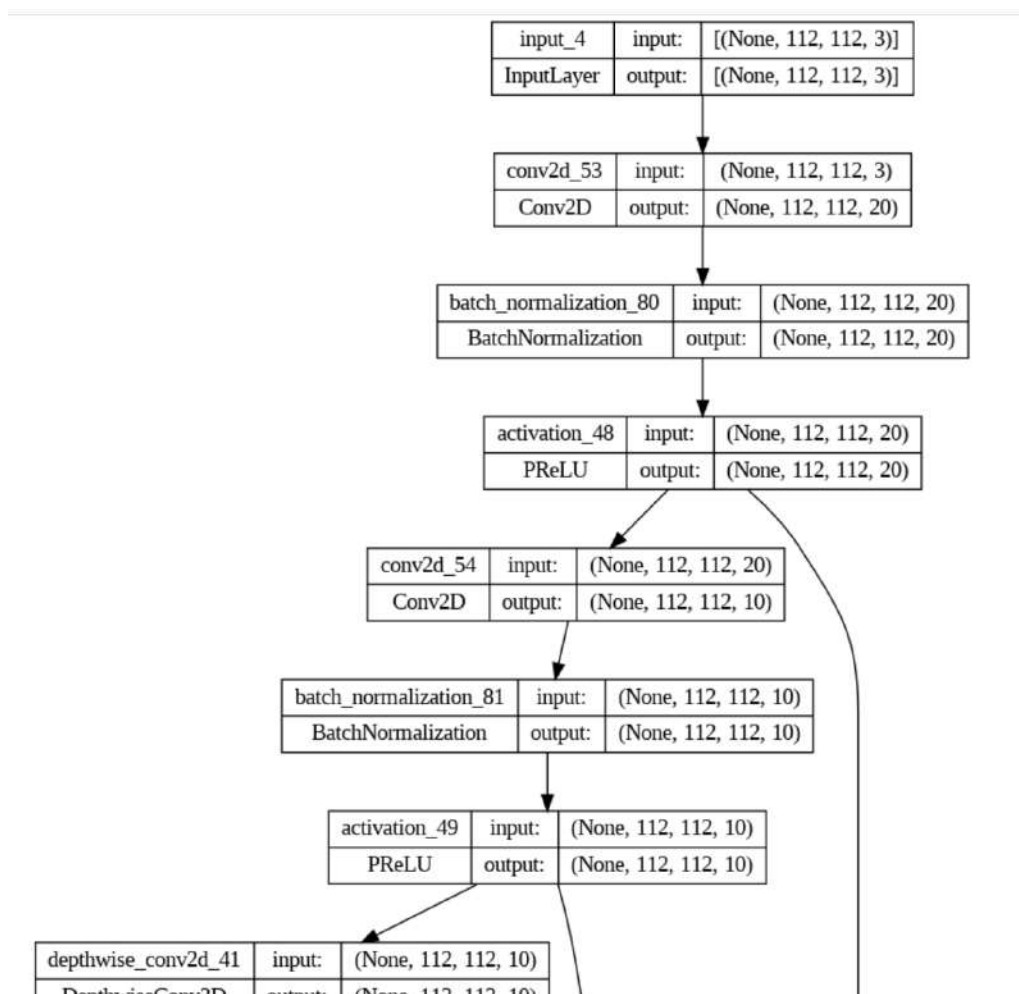
Setelah algoritma menghasilkan model yang berbeda-beda di tiap iterasinya, model dengan kombinasi akurasi yang tinggi dan loss yang rendah kemudian diuji berulang kali pada data video yang telah diambil menggunakan CCTV pada lingkungan sekolah SD Telkom Makassar.

Dari total 5600 model yang dihasilkan, didapatkan bahwa model terbaik memiliki parameter dan hyperparameter berikut:

- Struktur dasar model: GhostFaceNets V1
- Dimensi embedding_layer: 512
- Learning rate: 0.009240999817848206
- Fungsi aktivasi: PReLU
- Frekuensi evaluasi: 1
- Jumlah layer terbentuk: 219
- Akurasi model: 0.9730
- Loss model: 15.5367

Setelah model terbaik didapatkan, model tersebut dapat digunakan berulang kali untuk melakukan deteksi dan rekognisi wajah pada gambar maupun video lainnya, pada penelitian ini model digunakan untuk mendeteksi dan mengelai wajah pada video berdasarkan nama-nama identitas yang disimpan dalam model. Hasil deteksi dan pengelanaan kemudian disajikan dalam tabel-tabel terpisah yang nantinya akan dianalisa. Gambar 4.8 berikut merupakan visualisasi model terbaik pada layer-layer awal. Setiap layer pada model Ghost Face Net merupakan komponen vital dalam

arsitektur jaringan saraf tiruan yang kompleks ini. Dengan menggunakan pendekatan yang mendalam terhadap pengenalan wajah, Ghost Face Net memanfaatkan sejumlah lapisan yang dirancang dengan cermat untuk mengatasi tantangan dalam pengenalan wajah, termasuk variasi pose, pencahayaan yang berbeda, dan ekspresi wajah yang kompleks.



Gambar 4. 8 Visualisasi tujuh layer awal pada mode yang terbentuk

Lapisan pertama dari Ghost Face Net adalah lapisan konvolusi, yang bertanggung jawab untuk mengekstraksi fitur-fitur penting dari gambar wajah. Lapisan konvolusi ini bekerja dengan memindai gambar menggunakan filter konvolusi untuk mengidentifikasi pola-pola visual yang signifikan. Dengan melakukan ini, lapisan konvolusi membantu model memahami detail-detail halus yang ada dalam gambar wajah, memungkinkan pengenalan yang lebih akurat.

Selanjutnya, Ghost Face Net menggunakan lapisan normalisasi batch untuk memastikan bahwa nilai-nilai yang digunakan oleh jaringan tetap berada dalam kisaran yang optimal. Normalisasi batch membantu mengurangi overfitting, mempercepat pelatihan, dan membuat model lebih stabil selama proses pembelajaran.

Lapisan-lapisan berikutnya termasuk lapisan aktivasi, seperti ReLU (Rectified Linear Unit), yang memperkenalkan non-linearitas ke dalam model. Non-linearitas sangat penting karena memungkinkan model untuk memahami hubungan yang kompleks antara fitur-fitur yang diekstraksi dari gambar wajah. Dengan demikian, model dapat mengenali pola-pola yang lebih rumit dan abstrak dalam data.

Ghost Face Net juga memasukkan lapisan dropout, yang secara acak mematikan sejumlah unit (neuron) dalam jaringan selama proses pelatihan. Ini membantu mencegah overfitting dengan memaksa jaringan untuk mengandalkan lebih banyak pada fitur-fitur yang relevan dan mengurangi ketergantungan pada fitur-fitur spesifik dari data pelatihan.

Lapisan-lapisan terakhir Ghost Face Net termasuk lapisan pooling, yang mengurangi dimensi data dengan memilih nilai-nilai yang paling signifikan. Ini

membantu mengurangi beban komputasi pada jaringan dan menghasilkan representasi yang lebih terfokus dan relevan dari gambar wajah.

Secara keseluruhan, setiap layer pada Ghost Face Net bekerja sama untuk menciptakan model pengenalan wajah yang sangat akurat dan dapat diandalkan. Dengan menggunakan kombinasi canggih dari teknik-teknik ini, Ghost Face Net telah membuktikan diri sebagai salah satu model terdepan dalam pengenalan wajah, membuka pintu bagi berbagai aplikasi termasuk keamanan, pengenalan emosi, dan interaksi manusia-mesin yang lebih baik.

Dalam pengujian menggunakan video, dalam 1 video yang telah diambil terdapat 3 skenario pengujian. Skenario-skenario tersebut bertujuan untuk mengetahui bagaimana performa sistem ketika wajah yang diketahui model hanya terdapat satu dalam satu adegan frame, bagaimana performa sistem ketika wajah yang diketahui model terdapat minimal 2 dalam satu adegan frame, dan terakhir bagaimana performa sistem ketika seluruh wajah yang diketahui model muncul secara bersamaan dalam satu adegan frame. Tabel 4.1, 4.2 masing-masing menampilkan hasil analisa bagaimana perfroma sistem dalam mengenali wajah. Nilai pada setiap sel pada tabel menunjukkan seberapa yakin model dalam memprediksi dan mercognisi wajah yang ada dalam frame.

Tabel 4. 1 Tabel Analisis Sistem Menggunakan Video CCTV Pada Skenario 1

		Hasil Pengenalan Wajah oleh Sistem (berdasarkan kelas / label)																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Non
Label / Kelas Asli dalam frame	1	0,8704																		
	2		0,8264																	
	3			0,9171																
	4				0,8047															
	5					0,8803														
	6						0,9141													
	7																			0,5249
	8																			0,5280
	9									0,8345										
	10										0,8645									
	11																			0,7025
	12																			0,5973
	13													0,8467						
	14														0,8845					
	15																			0,7077
	16																0,9102			
	17																	0,8934		
	18																			0,7178
	Non														0,8124					

Pada tabel 4.1. tersebut, setiap pasangan kolom-baris yang memiliki nilai menunjukkan bagaimana sebuah wajah dideteksi pada scenario 1. Pada scenario 1 ini, setiap orang yang telah dikenali oleh model akan melewati jalur pulang sekolah di SD Telkom Makassar dan masuk kedalam frame CCTV sendiri-sendiri. Selain 18 wajah yang telah dikenali model, terdapat juga beberapa orang dalam frame CCTV yang tidak dilakukan perekaman dan pengambilan data wajahnya, sehingga orang tersebut dianggap non_label atau wajah yang seharusnya tidak diketahui oleh model.

Namun pada saat pengujian, kelas 7,8,11,12,15, dan 18 hingga akhir scenario pertama tidak dikenali oleh sistem. Sementara kelas non_label beberapa kali dikenali sebagai wajah dengan label 14, namun ini terjadi ketika wajah kelas 14 yang asli

belum masuk kedalam frame. Gambar 4. 9 berikut ini menunjukkan beberapa contoh adegan pada scenario pengujian 1.



Gambar 4. 9 Beberapa contoh hasil pendeteksian oleh sistem pada scenario 1

Dalam video yang sama juga diambil data dengan scenario, setiap pasangan guru-anak masuk kedalam frames secara bersamaan (terdapat 2 orang dalam frame).

Tabel 4.2 berikut menunjukkan hasil analisis pengujian dengan scenario kedua ini.

Tabel 4. 2 Tabel Analisis Sistem Menggunakan Video CCTV Pada Skenario 2

			Hasil Pengenalan Wajah oleh Sisitem (berdasarkan kelas / labele)																					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Non	Non		
Label / Kelas Asli dalam frame	1	9	0,8299								0,9280													
	2	10																			0,7826	0,7571		
	3	11			0,9007																0,7377			
	4	12				0,9000																	0,6940	
	5	14					0,8599									0,9159								
	6	13						0,9542															0,5858	
	7	15															0,8165				0,7033			
	8	16																			0,7212	0,5974		
	17	18																			0,7264	0,6394		

Dalam pengujian scenario kedua, terjadi penurunan performa dari GhosFaceNet, beberapa wajah yang awalnya berhasil di deteksi pada scenario pertama menjadi teridentifikasi sebagai orang yang tidak dikenali. Hal ini dikarenakan

semakin banyak objek wajah dalam frame, maka interferensi antara fitur wajah akan semakin sulit diinterpretasikan, algoritma akan lebih banyak menerjemahkan fitur wajah dan menghitung kemungkinan fitur terdekat dengan data wajah yang telah disimpan. Selain itu, proses deteksi dan rekognisi pengenalan wajah pada algoritma GhostFaceNet berjalan sebagaimana umumnya algoritma deteksi dan rekognisi wajah yang hanya dapat memproses data dari frame ke frame secara terpisah. Hal ini tentu menyebabkan terjadinya proses paralelisasi yang dapat membebani perangkat. Untuk itu, hasil pada laporan kemajuan ini masih perlu ditingkatkan hingga pada tingkat algoritma dapat tetap bekerja optimal bahkan jika terdapat 50 wajah dalam satu frame (hal ini merujuk pada luasan wilayah tunggu di gerbang sekolah). Gambar 4.10 berikut merupakan beberapa contoh hasil pendeteksian pada scenario kedua.



Gambar 4. 10 Beberapa contoh hasil pendeteksian oleh sistem pada scenario 2

Pada gambar terlihat bahwa bounding box tidak tergambar tepat pada pemilik wajah. Hal ini dikarenakan proses menggambar bounding box lebih lambat dari kecepatan perpindahan frame. Hal ini dapat diatasi dengan cara memberikan interval waktu pada algoritma untuk melakukan deteksi, rekognisi, dan menggambar bounding box, namun ada penelitian ini, interval waktu diatur menjadi 1 frame/deteksi dan

rekognisi, hal ini untuk melihat secara detail kemampuan algoritma dalam mendeteksi dan melakukan rekognisi. Jika interval dinaikkan, maka bounding box dapat tergambar dengan lebih baik, namun beberapa informasi terkait wajah pada frame sebelumnya menjadi tidak tersimpan.

Pada scenario terakhir, seluruh wajah yang terdaftar masuk secara bersamaan dengan tetap berpasang-pasangan antara murid dan guru, hal ini dikarenakan scenario ini akan menjadi bahan evaluasi juga pada tahap pengujian fitur peringatan dini berbasis WhatsApp. Hasil dari scenario ini menunjukkan hanya 4 wajah dari total 18 wajah yang masuk yang berhasil dikenali dengan baik oleh sistem. Hal ini dikarenakan peumpukan fitur wajah sehingga algoritma mengeluarkan skor keyakinan pada tiap wajah dengan nilai dibawah batas minimum yaitu 0.8. Hal ini dibuktikan pada Gambar 4.11 yang mana pada scenario ini hanya wajah yang berada di depan yang berhasil di deteksi dan di kenali.



Gambar 4. 11 Hasil pendeteksian oleh sistem pada scenario 3

Problem kegagalan pendeteksian dengan hanya menggunakan GhostFaceNet secara teoritis dapat ditangani dengan mengintegrasikan algoritma tracking yang akan menyimpan data wajah dan skor hasil recognisinya pada tiap frame, sehingga meskipun posisi wajah berubah bounding box akan mengikuti serta skor pada GhostFaceNet akan terus diupdate sehingga mengurangi kesalahan pengelanaan wajah.

4.6. Tahap Integrasi GhostFaceNet – DeepSORT

DeepSORT merupakan algoritma tracking atau pelacakan pada video. Penting untuk dicatat bahwa DeepSORT adalah komponen dari *pipeline* pelacakan yang lebih besar dan bekerja bersama dengan model deteksi objek (untuk menyediakan deteksi, dalam hal ini mendeteksi dan mengenali wajah) dan model ekstraksi fitur tampilan (untuk menyediakan fitur tampilan). Data masukan ke DeepSORT biasanya berasal dari komponen upstream dari algoritma pendeteksian, dan peran DeepSORT adalah mengaitkan deteksi lintas atau antar frame dan memelihara trek objek.

Pada penelitian ini prose integrasi kedua algoritma dapat digambarkan dengan pseudocode berikut:

```
CLASS OutputClassificationData:
    CONSTRUCTOR() :
        bbs = None
        classes = None
        category_index = None
        score_thresh = None
        scores = None
        image_np = None
```

```

FUNCTION      recognize_faces_and_track(det,      face_model,      image_classes,
embeddings, tracker, frame, output_data, dist_threshold=0.6):
    rec_dist, rec_class, bbs, ccs, rec_idx = image_recognize(image_classes,
embeddings, det, face_model, frame)
    updated_labels = []

    label_map, num_classes = get_label_map(image_classes)
    class_ids = []
    FOR EACH dist, label IN zip(rec_dist, rec_class):
        IF dist < dist_threshold THEN
            label = "Unknown"
        END IF
        updated_labels.append(label)

        class_id = FIRST ITEM IN label_map WHERE label_map[item]["name"] IS
EQUAL TO label
        class_ids.append(class_id["id"])
    END FOR

    detection_list = []
    boxes_norm = []
    FOR EACH bb, label IN zip(bbs, updated_labels):
        left, up, right, down = bb
        h, w = frame.shape[:2]
        boxes_norm.append([up/h, left/w, down/h, right/w])
        detection_list.append([left, up, right - left, down - up, label])

    output_data.bbs = CONVERT TO NUMPY ARRAY boxes_norm
    output_data.classes = CONVERT TO NUMPY ARRAY class_ids
    output_data.category_index = label_map
    output_data.score_thresh = dist_threshold
    output_data.scores = CONVERT TO NUMPY ARRAY rec_dist
    output_data.image_np = frame
    output_data = tracker.run(output_data, frame)

    PRINT "bbs", output_data.bbs
    PRINT "scores", output_data.scores
    RETURN output_data

```

```

FUNCTION    video_recognize_and_track(image_classes,    embeddings,    det,
face_model, tracker, video_source=0, frames_per_detect=5, dist_thresh=0.6,
nms_max_overlap = 1.0):
    cap = OPEN VIDEO CAPTURE WITH video_source
    cur_frame_idx = 0

    output_data = CREATE NEW OutputClassificationData OBJECT
    WHILE True:
        grabbed, frame = cap.read()
        IF NOT grabbed THEN
            BREAK
        END IF

        IF cur_frame_idx MOD frames_per_detect IS EQUAL TO 0 THEN
            output_data    =    recognize_faces_and_track(det,    face_model,
image_classes, embeddings, tracker, frame, output_data, dist_thresh)
            cur_frame_idx = 0
        END IF

        key = WAIT KEY INPUT WITH DELAY 1 AND APPLY 0xFF MASK
        IF key IS EQUAL TO ASCII VALUE OF "s" THEN
            SAVE frame AS "{}.jpg".format(cur_frame_idx)
        END IF

        IF key IS EQUAL TO ASCII VALUE OF "q" THEN
            BREAK
        END IF

        frame    =    helper.drawDetectedBBs(frame.copy(),    output_data,
score_thresh=0.1)
        SHOW frame
        cur_frame_idx += 1

    RELEASE VIDEO CAPTURE
    CLOSE ALL OPEN WINDOWS

```

Dalam pseudo code tersebut, hasil dari deteksi dan rekognisi wajah untuk tiap frame akan masuk kedalam algoritma DeepSORT dalam bentuk dua informasi

penting; index atau urutan terjadinya deteksi dan rekognisi pertama kali serta informasi posisi objek hasil rekognisi. Untuk tiap frame ke- i terjadinya deteksi dan rekognisi kemudian akan dibandingkan dengan frame setelahnya pada urutan ke- $i+1$, apabila terdapat objek yang hasil rekognisinya pada frame yang berbeda tersebut dan posisinya tidak lebih dari 20 pixel maka bounding box dari objek tersebut akan ditandai sebagai objek yang sama dan diberikan ID hasil rekognisi. ID kemudian akan terus dipertahankan sehingga objek yang sama pada beberapa frame akan sama hingga objek tersebut menghilang dari frame. Hal ini menjadikan perubahan posisi objek pada tiap frame akan di tracking sehingga proses deteksi dan rekognisi akan menjadi lebih akurat dan presisi. Hasil dari integrasi algoritma GhostfaceNet dan DeepSORT pada skenario pertama kemudian dapat diamati pada gambar 4.12 berikut:

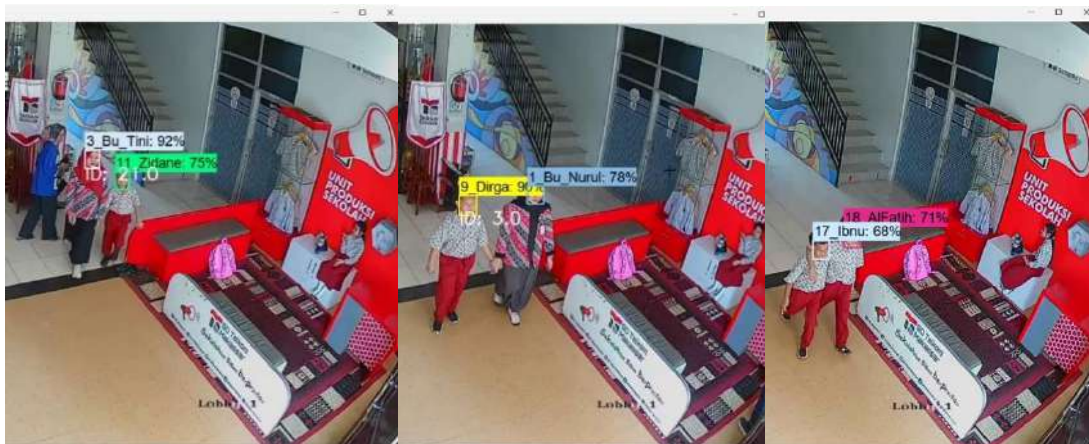


Gambar 4. 12 Beberapa contoh hasil deteksi + rekognisi + tracking oleh sistem pada skenario 1

Jika merujuk pada gambar 4.12 dan gambar 4.9, hasil pembentukan bounding box lebih optimal dikarenakan posisi bounding box yang berfungsi sebagai penanda lokasi wajah tergambar tepat di wajah tiap objek. Hal ini tentu menjadi hal peningkatan yang signifikan mengingat tujuan akhir dari sistem adalah mendeteksi jika siswa keluar

dari wilayah sekolah bersama orang asing, maka posisi akurat bounding box akan menjadi acuan saat siswa masuk kedalam wilayah ROI yang nantinya akan dibuat.

Sebagaimana hasil pengujian pada scenario kedua dan ketiga sebelum algoritma diintegrasikan, hasil pengujian setelah diintegrasikan juga mengalami sedikit penurunan performa dari sisi ketepatan hasil rekognisi namun hal tersebut tidak menjadi sebuah masalah dikarenakan sebelum objek keluar dari wilayah frame hasil akhir dari tracking menunjukkan hasil yang tepat. Gambar 4.12 berikut merupakan contoh hasil rekognisi yang disertai tracking pada scenario pengujian kedua.



Gambar 4. 13 Beberapa contoh hasil deteksi + rekognisi + tracking oleh sistem pada scenario 2

Merujuk pada perbandingan hasil pada gambar 4.13 dan 4.10, terlihat bahwa hasil rekognisi dan pengenalan wajah menjadi lebih optimal dengan beberapa penurunan skor *distance confidence*, dimana skor ini menunjukkan seberapa percaya model terhadap hasil rekognisi dan tracking nya sendiri. Penurunan skor ini disebabkan oleh penambahan jumlah objek dalam frame menyebabkan kompleksitas spasial sehingga objek yang masuk lebih awal dalam frame akan memiliki prioritas

Figure 10 displays four screenshots of a video surveillance system, showing people in a room. Each screenshot includes bounding boxes and text labels indicating the system's identification of individuals and their confidence scores. The room features red walls, a staircase, and various posters.

- Top Left Screenshot:**
 - 12 Ayunda: 70%
 - 2 Bu. Suryani: 67%
 - 9 Dirga: 75%
 - Nurul: 78%
 - 5 Pak. Fikar: 71%
 - 14 Tarita: 88%
 - ID: 150
- Top Right Screenshot:**
 - Urutawan: 40%
 - 5 Pak. Fikar: 72%
 - 13 Aglah: 71%
 - Idht: 90%
 - ID: 70
- Bottom Left Screenshot:**
 - 10 Fatm: 75%
 - 2 Bu. Suryani: 67%
 - 10 Fatm: 72%
 - 1 Bu. Nurul: 58%
 - 9 Dirga: 68%
- Bottom Right Screenshot:**
 - 12 Ayunda: 63%
 - 3 Bu. Tini: 91%
 - 15 Es. ps: 70%

4.7. Pembuatan ROI Wilayah Sekolah

Untuk mendefinisikan ROI, kita perlu tahu batas-batas yang menjadi acuan dalam video atau gambar untuk menentukan seseorang keluar atau masuk ke wilayah

sekolah. Jika hasil deteksi, rekognisi dan tracking masih berada di wilayah ROI maka bisa dipastikan seseorang masih dalam wilayah sekolah. Untuk mendefenisikan hal tersebut, maka kita perlu tahu posisi pixel batas tersebut terhadap batas di tempat pengambilan data.

Pada penelitian ini, ROI dibangun dengan menggunakan opencv dengan bentuk persegi panjang dan secara langsung tampil pada video frame. Adapun baris code dalam membuat ROI pada penelitian ini adalah sebagai berikut:

```
# Draw a fixed ROI box at a specific location
roi_x = 500 # X-coordinate of the top-left corner of the ROI
roi_y = 500 # Y-coordinate of the top-left corner of the ROI
roi_width = 930 # Width of the ROI
roi_height = 350 # Height of the ROI
cv2.rectangle(frame, (roi_x, roi_y), (roi_x + roi_width, roi_y
+ roi_height), (0, 255, 0), 2)
```

Pada penelitian ini, ukuran frame video yang terproses dalam algoritma adalah 1920 x 1080 piksel yang kemudian ROI didefenisikan dengan ukuran 930x350 piksel yang titik pusat ROI terletak pada pixel ke-500. Pembuatan ROI ini kemudian harus dilakukan penyesuaian ketika sistem akan digunakan pada device kamera yang baru. Adapun hasil pembentukan ROI dapat dilihat pada Gambar 4.15 berikut:



Gambar 4. 15 Hasil pembuatan ROI pada sistem

Pemilihan lokasi ROI sebagaimana yang tampak pada gambar 4.15 dikarenakan pada posisi tersebut model sudah cukup optimal dalam melakukan pengenalan wajah, serta pada titik tersebut merupakan pintu keluar siswa dari wilayah sekolah, sehingga proses untuk menentukan apakah seorang siswa keluar dari wilayah sekolah bersama orang asing akan dimulai ketika terdapat siswa masuk kedalam ROI.

4.8. Database Management

Dalam tahap awal pembuatan aplikasi, model tidak dapat mengenali keterkaitan hubungan antara satu siswa dengan guru. Model hanya mengetahui identitas seperti nama yang telah didaftarkan pada saat proses pelatihan. Namun, info lanjut mengenai

seorang siswa terasosiasi dengan guru atau orang tua yang mana diperlukan mekanisme lanjutan dalam database untuk saling menentukan hubungan.

Selain itu, database juga berfungsi untuk mengenali status seseorang berdasarkan nama yang telah diberikan oleh model. Didalam database, nama siswa akan disimpan pada kolom siswa, nama seorang guru akan disimpan pada kolom guru, dan begitu seterusnya. Selain itu, manajemen databse ini juga berfungsi untuk mengelompokkan model dan identitas berdasarkan sekolah. Sehingga nantinya setiap sekolah akan memiliki model deteksi dan rekognisi masing-masing.

File yang menyimpan nama siswa , pendamping siswa, guru, serta nomor kontak yang akan dikirimkan pesan notifikasi akan disimpan dalam format .Json dengan bentuk sebagai berikut:

```
[
  {
    "siswa": "9_Dirga",
    "pendamping": ["1_Bu_Nurul"],
    "no_hp": "08123"
  },
  {
    "siswa": "10_Fatan",
    "pendamping": ["2_Bu_Suryani"],
    "no_hp": "08123"
  },
]
```

Ketika sistem mendeteksi wajah, maka setiap nama yang di deteksi kemudian akan di cek didalam database, apakah personal tersebut termasuk siswa atau pendamping. Jika yang masuk kedalam ROI adalah siswa maka sistem akan mendeteksi apakah terdapat pendamping yang berasosiasi dengan siswa tersebut berdasarkan file .json yang telah dibuat.

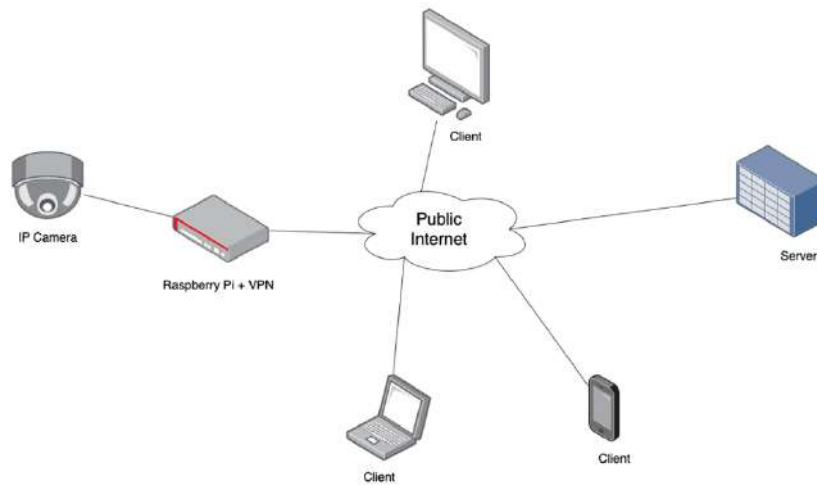
Dalam rencana jangka panjang dalam menerapkan sistem ini secara menyeluruh di Kota Makassar nantinya terdapat dua opsi terkait storage dan penyimpanan data dan sistem; Penyimpanan terpusat yang ditangani oleh pemerintah kota makassar, dan Penyimpanan terdistribusi dimana setiap sekolah menyimpan data masing-masing namun sistem yang dibangun ditangani oleh pemerintah kota makassar. Pada opsi pertama, setiap sekolah akan menyetorkan data wajah dan identitas tiap individu dan server terpusat milik pemerintah kemudian akan melakukan proses pelatihan model yang dikhususkan untuk tiap sekolah dan kemudian keamanan dan privasi data akan diserahkan kepada tenaga ahli pada lingkup pemerintah kota makassar. Pada opsi yang kedua, setiap sekolah akan melakukan request kepada pemerintah kota untuk dibuatkan model rekognisi tanpa mengirimkan data mereka masing-masing, namun opsi kedua ini mewajibkan tiap sekolah untuk memiliki tenaga ahli masing-masing yang bertanggung jawab atas keamanan server dan privasi setiap individu. Dari dua opsi tersebut, opsi menjadi pilihan terbaik dimana kemungkinan kebutuhan biaya dapat lebih efisien dikarenakan pihak pemerintah kota dapat lebih bijak dalam mengalokasikan dana terkait kebutuhan infrastruktur dibandingkan hal tersebut dibebankan untuk tiap sekolah.

Database management pada sistem yang dibangun juga sudah menyediakan fleksibilitas, dimana setiap siswa dapat diasosiasikan dengan lebih dari 1 pendamping dan nomor WhatsApp. Hal ini dapat diamanti pada contoh file .Json sebelumnya, dimana setiap nama dan nomor WhatsApp dapat secara langsung ditambahkan dalam file tersebut berdasarkan penamaan saat pelatihan model.

4.9. Menghubungkan Kamera IP dan RaspberryPi

Menghubungkan Kamera IP ke Raspberry Pi adalah langkah penting dalam mengintegrasikan sistem pemantauan. Prosesnya melibatkan pemilihan kamera IP yang sesuai, konfigurasi jaringan, identifikasi alamat IP kamera, instalasi perangkat lunak pendukung di Raspberry Pi, konfigurasi kamera pada perangkat, dan pengujian koneksi. Setelah berhasil terhubung, kamera IP dapat diintegrasikan dalam Raspberry Pi, memungkinkan kita dapat mengakses sistem dengan perangkat yang lain secara terpisah.

Tujuan penggunaan Raspberry Pi tidak hanya untuk aksesibilitas, melainkan juga untuk keamanan. Data yang dikirim langsung tanpa melalui Raspberry Pi akan sangat rentan terhadap serangan dari luar. Oleh karena itu Raspberrpi Pi juga berfungsi untuk mengirimkan data video melalui Virtual Private Network (VPN). Ketika Anda terhubung ke VPN, semua data yang dikirim melalui koneksi dienkripsi. Ini berarti bahwa informasi yang dikirim, seperti situs web yang dikunjungi, kata sandi yang dimasukkan, dan data berupa video, diubah menjadi kode yang sangat sulit untuk diuraikan. Enkripsi ini melindungi data dari pihak ketiga yang mencoba mengintip atau mencuri informasi. Berikut skema sistem secara keseluruhan dalam jaringan.



Gambar 4. 16 Skema jaringan pada sistem

4.10. Pengiriman Video ke Server melalui Koneksi Socket

Pengiriman video ke server melalui koneksi socket adalah proses mentransfer aliran video secara real-time dari Raspberry Pi ke server yang terletak di jaringan atau internet. Teknik ini sangat berguna dalam berbagai aplikasi, seperti video streaming, pemantauan keamanan, atau komunikasi jarak jauh.

Dalam konteks ini, penggunaan socket memungkinkan komunikasi langsung antara perangkat pengirim (misalnya, kamera) dan server. Video yang dihasilkan oleh perangkat pengirim dipecah menjadi paket-paket kecil, yang kemudian dikirimkan ke server melalui saluran koneksi socket yang terbuka. Proses ini memungkinkan data video untuk dikirim dalam *real-time* dan dapat diurai oleh server untuk kemudian diproses atau disalurkan kepada pengguna. Selama pengiriman video, menggunakan protokol komunikasi RTSP (Real-Time Streaming Protocol).

Dalam server memuat kode rekognisi dari algoritma GhostFaceNets. Video yang telah dikirim dari RaspberryPi yang dikirim melalui Socket kemudian di Rekognisi dalam server. Hasil dari rekognisi dalam server kemudian dikirim dalam bentuk pesan melalui Whatsapp.

Namun, dalam penerapannya pengiriman data berupa video real-time bergantung dengan kecepatan koneksi internet masing-masing. Dalam penelitian ini *delay* yang didapatkan saat mengirim video mencapai 10 detik bahkan dapat lebih tergantung koneksi internet. Sangat tidak ideal bagi system yang mengutamakan peringatan dini.

Socket sendiri bukanlah alat keamanan, tetapi merupakan mekanisme dasar yang digunakan dalam komunikasi jaringan. Keamanan dalam komunikasi socket sangat bergantung pada bagaimana socket digunakan, konfigurasi, dan langkah-langkah yang diambil untuk melindungi komunikasi. Untuk penelitian lebih lanjut perlu meneliti bagaimana keamanan data saat mengirim data dengan video secara cepat dan aman. Berikut adalah beberapa pertimbangan terkait dengan keamanan socket:

1. Enkripsi: Untuk meningkatkan keamanan komunikasi, Anda dapat mengenkripsi data yang dikirim melalui socket. Ini dapat dicapai dengan menggunakan protokol seperti HTTPS untuk koneksi web atau Secure Socket Layer (SSL) untuk koneksi umum. Dengan enkripsi, data akan diacak sehingga hanya pihak yang memiliki kunci enkripsi yang sesuai yang dapat membacanya.
2. Otentikasi: Pastikan bahwa pihak yang terhubung ke socket adalah entitas yang sah. Ini dapat dilakukan melalui mekanisme otentikasi, seperti

penggunaan kata sandi atau sertifikat digital. Dengan otentikasi yang kuat, Anda dapat memastikan bahwa hanya pihak yang memiliki izin yang tepat yang dapat terhubung.

3. Firewall: Menggunakan firewall yang benar adalah penting untuk melindungi socket dari akses yang tidak diinginkan. Firewalls dapat membatasi akses ke socket, mengontrol lalu lintas jaringan, dan mencegah serangan dari pihak yang tidak sah.
4. Pembaruan dan Keamanan Sistem Operasi: Pastikan sistem operasi yang menjalankan socket selalu diperbarui dengan perbaikan keamanan terbaru. Ini membantu mengatasi kerentanan yang mungkin ditemukan dan dieksploitasi oleh penyerang.
5. Kendali Akses: Menerapkan kendali akses yang ketat pada server socket adalah penting. Ini mencakup memastikan bahwa hanya aplikasi atau entitas yang sah yang memiliki akses ke socket dan data yang dibagikan melalui socket.
6. Monitoring dan Logging: Memantau aktivitas socket dan mencatat log adalah langkah yang baik untuk mendeteksi kejanggaran atau upaya peretasan. Dengan log yang tepat, Anda dapat mengidentifikasi ancaman keamanan lebih cepat.
7. Pembaruan dan Pengujian Aplikasi: Pastikan bahwa aplikasi yang menggunakan socket diperbarui dan diuji secara teratur untuk mengidentifikasi dan memperbaiki kerentanan keamanan.

Keselamatan socket bergantung pada cara socket digunakan, diimplementasikan, dan dilindungi. Dengan tindakan yang benar, socket dapat digunakan secara aman dalam berbagai aplikasi jaringan.

4.11. Notifikasi WhatsApp

Notifikasi ke Orang Tua melalui pesan whatsapp jika hasil deteksi mengatakan jika anak tersebut Bersama dengan orang yang tidak biasanya. Pengiriman notifikasi kepada orang tua siswa menggunakan *pywhatkit* adalah sebuah pendekatan yang memanfaatkan perpustakaan Python. Proses dimulai dengan pengumpulan informasi kontak orang tua atau wali siswa, seperti nomor telepon atau rincian kontak WhatsApp. Kemudian, dengan menggunakan Python dan perpustakaan *pywhatkit*, sekolah dapat mengirimkan pesan notifikasi yang dapat dipersonalisasi kepada orang tua siswa secara otomatis.

Dalam system ini pesan yang dikirim berupa teks peringatan dan gambar saat system mendeteksi anak sedang bersama dengan orang tidak dikenal. Dalam pesan tersebut secara otomatis mengambil nama anak, pendamping, waktu, dan nomor orang tua pendamping dari database. Selain mengirim, system ini juga menyimpan hasil histori notifikasi.

Tujuan penggunaan *Pywhatkit* ini hanya untuk prototipe system. Tentu ada beberapa kekurangan dari *pywhatkit* ini seperti sangat bergantung pada User Interface pengguna. Sangat direkomendasikan menggunakan layanan berbayar pengirim pesan WhatsApp seperti Twilio untuk performa yang lebih baik.

4.12. Analisis Kebutuhan Infrastruktur

Sebagaimana yang telah diketahui bahwa penelitian ini merupakan upaya membangun purwarupa sistem yang dapat mendeteksi apabila seorang anak keluar dari wilayah sekolah bersama orang tidak dikenal maupun keluar dari wilayah sekolah tanpa pengawasan. Untuk mengimplementasikan sistem secara menyeluruh, hal-hal yang perlu diperhatikan adalah sebagai berikut:

4.12.1. Kebutuhan Storage

Dalam penelitian ini, satu model GhostFaceNet yang dihasilkan saat proses pelatihan adalah berkisar pada 30 – 170 Megabyte. Sedangkan ukuran gambar yang belum memasuki proses pre-processing dan augmentasi adalah 20 KiloByte – 1 Megabyte tergantung jenis kamera yang digunakan, dan setelah dilakukan pre-processing dan augmentasi seluruh data menjadi berukuran 5 – 10 KiloBytes. Adapun ukuran file keseluruhan file script code untuk kode ini adalah 74 MegaByte.

Dengan asumsi, setiap sekolah minimal memiliki 200 murid dengan masing-masing murid didaftarkan 2 pendamping, jumlah guru adalah 50 orang dan proses pelatihan dilakukan 10 kali percobaan yang masing-masing dilakukan 200 kali iterasi. Maka kebutuhan storage untuk menerapkan sistem pada satu sekolah adalah akumulasi dari: Jumlah total ukuran model + Total ukuran Gambar sebelum mengalami pre-processing + Total Ukuran gambar setelah mengalami pre-processing + Ukuran script Sistem + Free Space.

Setiap iterasi memproduksi satu model, maka kebutuhan storage untuk menyimpan model adalah 170 Megabyte x 200 (iterasi) x 10 (percobaan) =

340.000 MegaByte \approx 340 GigaByte. Adapun ukuran gambar asli yang belum melalui pre-processing dan augmentasi adalah 1 MegaBytes x {200 (anak) + 400 (pendamping) + 50 (guru)} x 30 Gambar = 19.500 MegaBytes \approx 19,5 GigaByte. Setelah data gambar melalui proses pre-processing dan augmentasi, total storage yang dibutuhkan adalah 10 KiloBytes x {200 (anak) + 400 (pendamping) + 50 (guru)} x 30 Gambar = 195.000 Kilobytes \approx 0,195 GigaByte.

Pada praktiknya, setiap sistem membutuhkan kekosongan (Free Space) sebesar 30% dari total storage agar sistem dapat berjalan optimal. Maka total kebutuhan storage yang dibutuhkan untuk membangun sistem ini untuk satu sekolah adalah 340 GigaByte (model) + 19,5 GigaByte (data awal) + 0,195 GigaByte (data pra pre-processing) + 0,074 GigaBytes (Script) + Free Space = 359,769 + 107,9307 (FreeSpace) = 467,6997 Gigabytes \approx 0,5 TeraBytes.

4.12.2. Kebutuhan Jenis Server

Dalam penelitian ini, device yang digunakan adalah Laptop dengan spesifikasi; Processor Intel Core i9-11900H, RAM 16GB, GPU NVIDIA GeForce RTX 3060. Spesifikasi ini dapat disetarakan dengan server dengan spesifikasi; Processor Intel Xeon W-1290P, RAM 32 GB, GPU NVIDIA GeForce RTX 3060.

4.12.3. Kebutuhan Perangkat Keras

Penelitian ini menggunakan IP Kamera yang telah disediakan oleh Dinas Komunikasi dan Informatika (Kominfo) Kota Makassar, dengan spesifikasi

2MP. IP Kamera dikoneksikan dengan PoE Injector yang terhubung melalui kabel LAN. Kabel LAN juga terhubung ke Raspberry Pi 4 8GB.

4.12.4. Perkiraan Kebutuhan Biaya

Dalam memperkirakan biaya yang dibutuhkan untuk mengimplementasikan sistem ini untuk satu sekolah dengan penjelasan kebutuhan berdasarkan poin-poin sebelumnya, maka total biaya yang dibutuhkan untuk menerapkan sistem ini adalah berkisar Rp. 55.000.000 – Rp. 70.000.000 (lima puluh hingga tujuh puluh juta rupiah) bergantung pada merek server yang digunakan. Adapun biaya tambahan diluar biaya infrastruktur adalah biaya berlangganan pesan otomatis WhatsApp pada layanan Twillo adalah 78 rupiah per pesan, sehingga dengan asumsi terburuk jika terjadi 100 kasus anak sekolah keluar bersama orang tidak dikenal, maka perharinya biaya peringatan adalah 7800 rupiah/hari. Hal ini tentu merupakan biaya yang sangat murah untuk keselamatan anak-anak di kota makassar.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil analisis terhadap pengujian Optimalisasi Algoritma Computer Vision Terhadap Kondisi Lingkungan Kota Makassar Untuk Mendeteksi Orang Tidak Dikenal Berbasis Video CCTV dapat disimpulkan bahwa:

1. Dalam upaya megembangkan sistem tracking multi-wajah dengan camera cctv diwilayah sekolah di kota makassar secara real-time dapat digunakan algoritma GhostFaceNetet sebagai algoritma dasar untuk mengenali wajah dan kemudian diintegrasikan dengan DeepSORT sebagai algoritma tracking.
2. Dalam upaya membangun sistem peringatan dini keamanan wilayah sekolah berbasis informasi cctv, penggunaan platform yang umum digunakan masyarakat dapat dimanfaatkan untuk meneruskan informasi secara real-time. Pada penelitian ini penggunaan aplikasi WhatsApp menjadi pilihan disebabkan umumnya platform ini digunakan sebagai media bertukar informasi yang privat dan tepat sasaran
3. Sistem yang dibangun pada penelitian ini sudah dapat digunakan dan di replikasi di wilayah sekolah di kota makassar, dengan akurasi pendeteksian yang berkisar pada 80 hingga 90 persen, serta delay yang didapatkan dalam mengirim informasi hanya 10 detik, yang mana ini dapat ditingkatkan dengan peningkatan infrastruktur dasar serta internet diwilayah kota makassar.
4. Penggunaan dan pemilihan algoritma pada tracking multi-wajah dengan camera cctv diwilayah sekolah di kota makassar secara real-time harus memperhatikan kecepatan dan ketepatan saat melakukan deteksi, hal ini mengingat banyaknya objek dan sekolah yang kedepannya dapat menggunakan sistem ini.
5. Peringatan dini untuk system membutuhkan Kamera IP CCTV yang terhubung dengan RaspberryPi. Data video CCTV tersebut dikirim ke server

dengan menggunakan Socket melalui jaringan VPN. Data yang telah dikirim ke server kemudian diolah dan hasil olahan tersebut dikirim melalui pesan masing-masing orang tua pada WhatsApp jika anak tersebut bersama dengan orang tidak dikenal.

6. Performa Algoritma GhostfaceNet yang diintegrasikan dengan DeepSort menunjukkan hasil yang cukup baik. Namun dalam proses pengiriman video sangat bergantung pada koneksi internet masing-masing. Selama penelitian *delay* yang tercatat selama proses pengiriman video melalui metode Socket mencapai 10 detik.
7. Ketersediaan infrastruktur merupakan kunci utama sistem yang dibangun dapat berjalan secara optimal.

5.2 Saran

Adapun saran untuk pengembangan penelitian selanjutnya adalah:

1. Dipaharapkan pada penelitian selanjutnya menambahkan metode kompresi pengiriman data untuk optimalisasi waktu pengiriman data berupa video.
2. Diharapkan pada penelitian selanjutnya juga memerhatikan keamanan pengiriman video. Metode pengiriman menggunakan Socket juga perlu dibarengi dengan algoritma enkripsi dan jenis keamanan lainnya.
3. Diharapkan untuk sistem ini menggunakan layanan berbayar untuk mengirim pesan WhatsApp seperti Twilio. Penggunaan layanan Pywhatkit sangat bergantung pada User Interface perangkat yang dikirim secara otomatis. Sehingga selama pengiriman pesan bisa saja terjadi error yang tidak diinginkan.
4. Sistem yang baik adalah sistem yang senantiasa diawasi dan dimaintenance, oleh karena itu penggunaan sistem ini pada tahap

pengimplementasian awal masih membutuhkan keterlibatan tenaga terlatih untuk mengevaluasi keakuratan pendeteksian. Hal ini untuk mengamati masalah-masalah baru yang belum coba diselesaikan pada penelitian ini.

Beberapa upaya yang dapat dilakukan untuk menyukseskan program jagai anakta' yang diinisiasi oleh pemerintah kota makassar, khususnya dalam lingkup sekolah ialah:

1. Pemasangan infrastruktur yang merata di persekolahan di kota makassar, khususnya terkait penyediaan CCTV yang memiliki spesifikasi yang seragam sehingga proses replikasi dan pemerataan teknologi yang telah dibangun dapat lebih menghemat waktu
2. Demi meringankan beban setiap sekolah yang ada di kota makassar, infrastruktur dasar yang cukup vital yang harus disediakan pemerintah kota makassar adalah server dengan spesifikasi yang telah disebutkan sebelumnya
3. Perlunya kerjasama dan kolaborasi antara institusi di wilayah kota makassar untuk menyukseskan program jagai anakta dengan media sistem yang telah dibangun
4. Salah satu keraguan yang kemungkinan akan muncul dalam pengimplementasian sistem ini adalah terkait keamanan data, maka dari itu pemerintah kota makassar sudah sepatutnya merekrut tenaga kermanan jaringan yang dapat bertanggung jawab atas keamana data

privasi hasil perekaman wajah. Sistem yang dibangun merupakan aplikasi yang telah siap digunakan, namun terkait keamanan berkaitan dengan server bukan proses pengkodean aplikasi.

5. Keterlibatan pihak kepolisian dalam menyukseskan program jagai anakta dapat lebih dioptimalkan dengan adanya sistem yang dibangun dengan cara mengikut sertakan pihak yang berwajib di wilayah kota makassar sebagai contact person saat sistem mendeteksi adanya anomaly.
6. Dasar dari sistem ini dapat dikembangkan untuk pengawasan di wilayah lain selain sekolah, namun perlu beberapa penyesuaian terkait logika kasustik yang ingin dilakukan *early warning*. Untuk itu, sistem ini dapat menjadi pilot project untuk pengawasan wilayah yang lebih praktis dan efisien.

Daftar Pustaka

- Adjabi, I., Ouahabi, A., Benzaoui, A., & Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. In *Electronics (Switzerland)* (Vol. 9, Issue 8, pp. 1–53). MDPI AG. doi: 10.3390/electronics9081188
- Ajit, A., Acharya, K., & Samanta, A. (2020, February 1). A Review of Convolutional Neural Networks. *International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE 2020*. doi: 10.1109/ic-ETITE47903.2020.049
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). *Understanding of a Convolutional Neural Network*.
- Arafah, M., Achmad, A., Indrabayu, & Areni, I. S. (2019). Face recognition system using Viola Jones, histograms of oriented gradients and multi-class support vector machine. *Journal of Physics: Conference Series*, 1341(4), 042005. <https://doi.org/10.1088/1742-6596/1341/4/042005>
- Avola, D., Cinque, L., Diko, A., Fagioli, A., Foresti, G. L., Mecca, A., Pannone, D., & Piciarelli, C. (2021). MS-Faster R-CNN: Multi-Stream Backbone for Improved Faster R-CNN Object Detection and Aerial Tracking from UAV Images. *Remote Sensing*, 13(9), 1670. <https://doi.org/10.3390/rs13091670>.
- Badan Pengembangan dan Pembinaan Bahasa, K. (2016b). Wajah. Kemdikbud. <https://kbbi.kemdikbud.go.id/entri/wajah>
- Chen, S., Berglund, E., Bigdeli, A., Sanderson, C., & Lovell, B. C. (2008). Experimental Analysis of Face Recognition on Still and CCTV Images. *2008 IEEE Fifth*

International Conference on Advanced Video and Signal Based Surveillance,
317–324. <https://doi.org/10.1109/AVSS.2008.15>

diskominfo. (2023, January 27). *Jagai Anakta', Pemkot Makassar & OCSEA UNICEF Kolaborasi Perkuat Perlindungan Anak*. DISKOMINFO.
<https://diskominfo.makassarkota.go.id/jagai-anakta-pemkot-makassar-ocsea-unicef-kolaborasi-perkuat-perlindungan-anak/>

Engineering, C., & Gables, C. (2017). Low Resolution Face Recognition in Surveillance Systems Using Discriminant Correlation Analysis. 912–917. <https://doi.org/10.1109/FG.2017.130>

Huang, T. S. (n.d.). *Computer Vision: Evolution and Promise*. 5.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2), 83–97.
<https://doi.org/10.1002/nav.3800020109>

Kusumanto, R. D., & Tompunu, A. N. (2016). PENGOLAHAN CITRA DIGITAL UNTUK MENDETEKSI OBYEK MENGGUNAKAN NORMALISASI RGB. January

Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2020). Face recognition systems: A survey. In *Sensors (Switzerland)* (Vol. 20, Issue 2). MDPI AG. doi: 10.3390/s20020342

Madenda, S. (2015). *Pengolahan Citra dan Video Digital* (A. M. Drajat (ed.)). Erlangga.

Mao, L., Sheng, F., & Zhang, T. (2019). Face Occlusion Recognition With Deep Learning in Security Framework for the IoT. *IEEE Access*, 7, 174531–174540.
<https://doi.org/10.1109/ACCESS.2019.2956980>

Mercioni, M. A., & Holban, S. (2020). The Most Used Activation Functions: Classic Versus Current. 15th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS, Suceava, Romania.

Putra, D. (2010). *Pengolahan Citra Digital*. Andi Offset.

R. E. KALMAN, R. E. K. (n.d.). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 11.

Rin. (n.d.). *Tercatat, 200 Kasus Kekerasan Anak di Makassar Tahun 2022*. Retrieved October 15, 2023, from <http://celebesmedia.id/celebes/artikel/1014120123/tercatat-200-kasus-kekerasan-anak-di-makassar-tahun-2022>

Shirai, Y. (1987). *Three-Dimensional Computer Vision*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-82429-6>

T. Sutoyo, S.Si, Edy Mulyanto, S.Si., M. ko. (2009). *Teori Pengolahan Citra Digital*

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 1, 1-511-1-518. <https://doi.org/10.1109/CVPR.2001.990517>

Visi Misi—BAPPEDA Makassar. (n.d.). Retrieved October 15, 2023, from <https://bappeda.makassarkota.go.id/profil/visi-misi>

Wibowo, E., Romika, D. P., & Ryana, M. C. (2017). Perancangan Video Profile Sebagai Penunjang Media Informasi Dan Promosi Pada Sma Cirta Islami Tangerang. *Seminar Nasional Teknologi Informasi Dan Multimedia 2017*, 61-66.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 3645-3649. <https://doi.org/10.1109/ICIP.2017.8296962>

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). WIDER FACE: A face detection benchmark. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016 - Decem, 5525–5533.
<https://doi.org/10.1109/CVPR.2016.596>

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23. doi: 10.1016/j.jii.2021.100224