

BAB 2 LANDASAN TEORI

Penelitian ini didasarkan pada sejumlah teori dan penelitian terdahulu yang relevan dengan pengembangan sistem presensi berbasis *face recognition*, yang dijabarkan sebagai berikut.

2.1 Penelitian Pendahulu

Tabel 2.1 menyajikan ringkasan penelitian terdahulu yang membahas penggunaan *Contrast Limited Adaptive Histogram Equalization (CLAHE)* dan model FaceNet dalam penerapan *face recognition*.

Tabel 2.1. Penelitian pendahulu

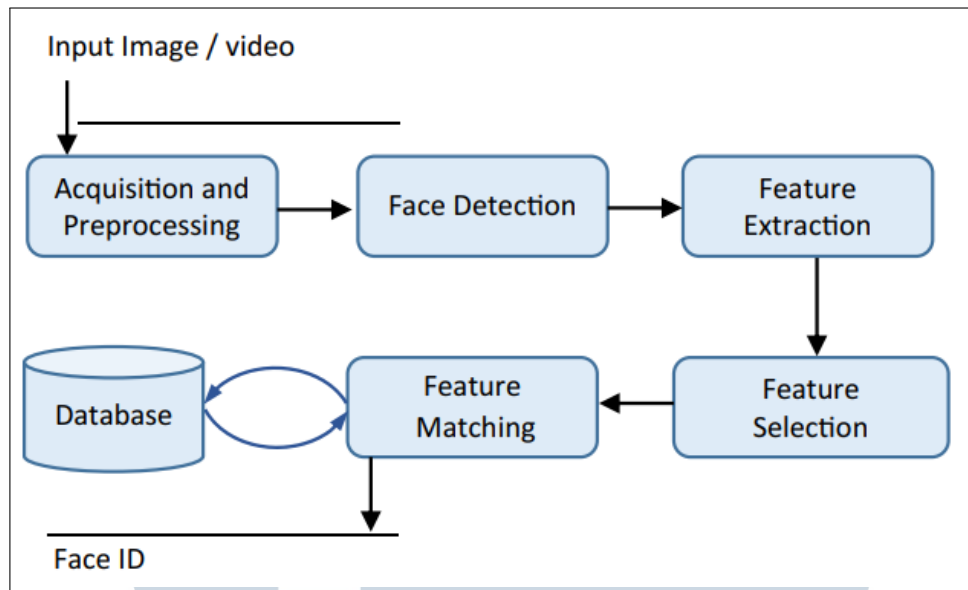
No	Aspek/Fitur Utama	Indira Hafiz (2023)	Kusnadi et al. (2021)	Malabi et al. (2024)	Sharmili et al. (2023)	Ajjour & Kurdy (2020)	Penelitian ini
1	Sistem presensi berbasis wajah	✓	×	✓	×	×	✓
2	Penggunaan model FaceNet	×	×	✓	×	✓	✓
3	Penerapan CLAHE untuk peningkatan kualitas citra	×	✓	×	✓	✓	✓
4	Implementasi pada sistem presensi kehadiran	✓	×	✓	×	×	✓
5	Implementasi sistem aplikasi (<i>mobile</i>)	✓	×	✓	×	×	✓
6	Integrasi dengan sistem informasi (HRIS)	×	×	×	×	×	✓

Dari tinjauan di atas, terlihat pola yang jelas bahwa penelitian sebelumnya umumnya hanya berfokus pada salah satu aspek, seperti penggunaan model FaceNet, penerapan metode CLAHE untuk peningkatan kualitas citra, atau implementasi sistem presensi berbasis wajah secara terpisah. Selain itu, belum terdapat penelitian yang secara komprehensif mengintegrasikan peningkatan kualitas citra menggunakan CLAHE dengan model FaceNet dalam fitur presensi pada sistem *Human Resource Information System (HRIS)* berbasis *mobile*. Hal ini menunjukkan adanya celah penelitian (*research gap*), yaitu belum optimalnya integrasi metode peningkatan citra dan model *deep learning* dalam konteks presensi terintegrasi HRIS. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan mengimplementasikan CLAHE dan FaceNet secara terintegrasi pada sistem presensi berbasis *mobile* yang terhubung dengan HRIS.

2.2 Face Recognition

Di era Revolusi Industri 4.0, perkembangan teknologi canggih, terutama kecerdasan buatan, telah mendorong kemajuan pesat dalam pengenalan wajah, yang menjadi salah satu tugas utama di bidang *computer vision* dengan berbagai penerapan potensial, mulai dari sistem keamanan dan presensi hingga layanan cerdas [19]. *Face Recognition* adalah teknologi yang digunakan untuk mengidentifikasi dan pengecekan identitas individu melalui analisis citra wajah [20]. Teknologi ini memanfaatkan ciri-ciri unik wajah (fitur wajah) seperti bentuk mata, hidung, mulut, dan kontur wajah untuk membedakan satu individu dengan yang lain. Selain itu, *face recognition* juga dapat diterapkan pada berbagai kondisi seperti pencahayaan, pose, dan ekspresi wajah.

Dalam hal ini, langkah fundamental sebelum melakukan *face recognition* adalah *face detection*, yang bertujuan menemukan wajah dalam gambar dengan teknik yang lebih cepat dan akurat [21].



Gambar 2.1. Sistem Pengenalan Wajah

Sumber: [21]

Terdapat beberapa tahapan utama dalam *face recognition* yang umumnya diterapkan, sebagaimana ditunjukkan pada Gambar 2.1.

1. Preprocessing Citra (*Image Preprocessing*).
Memperbaiki kualitas citra wajah supaya lebih konsisten, misalnya dengan *normalisasi*, *cropping*, *resizing*, atau *enhancement* seperti CLAHE untuk meningkatkan kontras sehingga menghasilkan citra mikrostruktur yang lebih jelas dan berkualitas tinggi [22].
2. Deteksi Wajah (*Face Detection*).
Memindai lokasi wajah dalam gambar atau video. Pada tahap ini sistem memisahkan wajah dari latar belakang.
3. Ekstraksi fitur (*Feature Extraction*).
Mengambil fitur wajah, seperti bentuk mata, hidung, mulut, atau kontur wajah.
4. Pencocokan (*Verification*).
Membandingkan fitur wajah yang diekstraksi dengan *database* untuk mengidentifikasi atau memverifikasi identitas individu.

2.3 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah salah satu jenis utama *neural networks* berbasis *multi-layer* yang digunakan dalam pengenalan dan klasifikasi gambar, dengan kemampuan untuk mengidentifikasi, mengenali, mengklasifikasikan, mendeteksi, serta melakukan segmentasi objek pada citra [23]. CNN dianggap sebagai *universal non-linear function approximator* karena digunakan untuk merepresentasikan informasi spasial dan memodelkan gambar dengan kemampuan tinggi dalam mengekstraksi fitur, sehingga dapat secara otomatis mempelajari hierarki fitur untuk klasifikasi [23]. Jaringan ini bekerja dengan memanfaatkan lapisan konvolusi, yaitu lapisan yang menerapkan operasi penyaringan (*filtering*) menggunakan *kernel* untuk mendeteksi pola lokal seperti tepi, tekstur, dan bentuk pada citra, sehingga fitur penting dapat diekstraksi secara bertahap [24]. Melalui proses ini, pola visual dapat dikenali secara lebih efektif.

CNN merupakan bagian dari bidang *Deep Learning* dalam kecerdasan buatan (*Artificial Intelligence/AI*), dan pada umumnya dilatih menggunakan pendekatan *supervised learning*, yaitu dengan memanfaatkan *dataset* berlabel agar jaringan dapat belajar mengenali pola dari *input* dan menghasilkan *output* sesuai target [25]. Pada penelitian ini, CNN yang digunakan adalah arsitektur FaceNet, yang berfungsi mengubah citra wajah menjadi representasi vektor berdimensi 128/512 (*embedding*) untuk kemudian dibandingkan menggunakan perhitungan jarak *Euclidean* [26].

Dalam CNN, proses konvolusi menghasilkan *feature map* dengan dimensi tertentu. Jika h dan w adalah tinggi dan lebar input, f adalah ukuran filter, s adalah *stride*, serta p adalah *padding*, maka dimensi keluaran dihitung dengan rumus [27].

$$h' = \left\lfloor \frac{h - f + 2p}{s} \right\rfloor + 1, \quad w' = \left\lfloor \frac{w - f + 2p}{s} \right\rfloor + 1 \quad (2.1)$$

Parameter *stride* (s) menyatakan langkah pergeseran *kernel* pada setiap operasi konvolusi. Jika $s = 1$, *kernel* bergeser satu piksel setiap kali perhitungan, sedangkan nilai s yang lebih besar menyebabkan *kernel* melompati beberapa piksel sekaligus. Nilai *stride* yang besar akan memperkecil dimensi *feature map* sehingga mengurangi kompleksitas komputasi, namun berpotensi menghilangkan sebagian detail spasial. Sebaliknya, nilai *stride* yang kecil mempertahankan lebih banyak informasi spasial tetapi menghasilkan ukuran keluaran yang lebih besar.

Ukuran *kernel* (f) dan *stride* (s) memiliki hubungan langsung terhadap dimensi keluaran dan cakupan informasi yang diproses. *Kernel* menentukan luas area lokal yang diamati dalam satu operasi konvolusi, sedangkan *stride* menentukan seberapa rapat area tersebut dipindai di seluruh citra.

Sementara itu, *receptive field* (RF) menunjukkan area input yang memengaruhi satu neuron pada *feature map*. Jika f adalah ukuran *kernel* dan s adalah *stride*, maka RF efektif untuk layer ke- n dapat ditulis sebagai [27].

$$RF^{(n)} = RF^{(n-1)} + (f_n - 1) \cdot \prod_{i=1}^{n-1} s_i \quad (2.2)$$

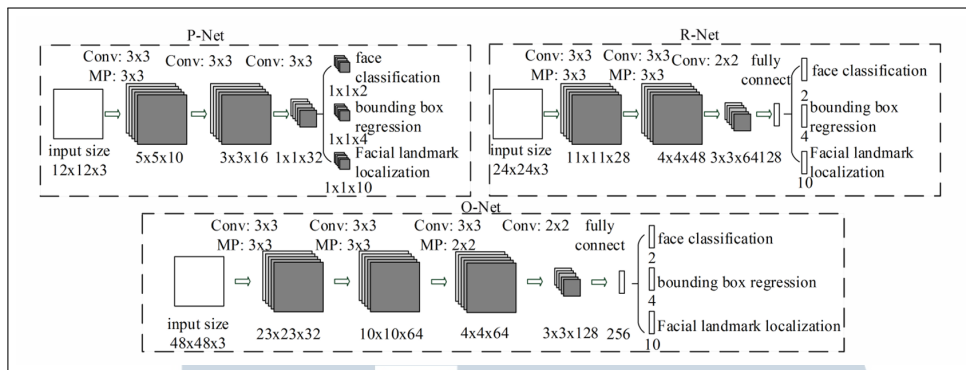
Keterangan variabel:

- (a) h, w = tinggi dan lebar citra input
- (b) f = ukuran filter/*kernel*
- (c) s = *stride* (langkah pergeseran *kernel*)
- (d) p = *padding* (penambahan piksel tepi)
- (e) RF = *receptive field*, area input yang berkontribusi pada satu neuron keluaran

Persamaan tersebut menunjukkan bahwa ukuran *kernel* berkontribusi langsung terhadap penambahan *receptive field*, sedangkan *stride* pada layer sebelumnya berperan sebagai faktor pengali yang memperluas cakupan informasi pada layer berikutnya. Perbandingan antar wajah dilakukan dengan menghitung jarak *Euclidean* antara vektor. Dua wajah dianggap sama apabila jaraknya berada di bawah ambang batas tertentu [26].

2.4 Multi-Task Cascaded Convolutional Neural Networks (MTCNN)

Multi-Task Cascaded Convolutional Neural Networks (MTCNN) adalah model hasil pengembangan dari arsitektur *Convolutional Neural Network* (CNN) yang terdiri atas tiga tahapan jaringan, yaitu P-Net, R-Net, dan O-Net, yang bekerja secara bertingkat melalui proses seleksi kandidat dan klasifikasi sehingga menghasilkan deteksi wajah yang lebih cepat dan efisien [28] [29].

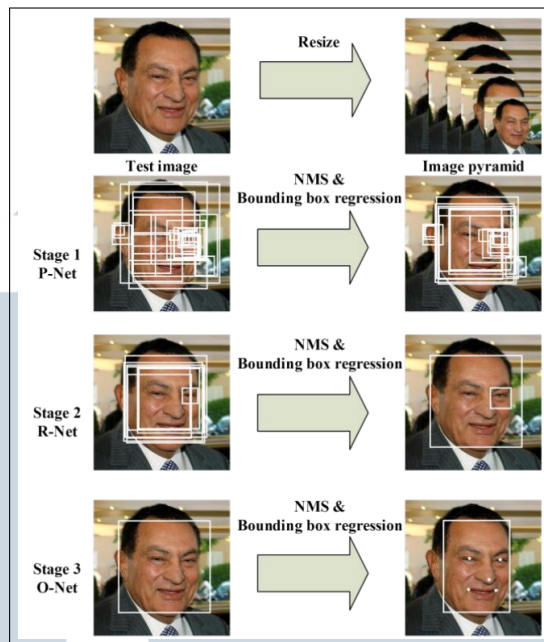


Gambar 2.2. Arsitektur *three-stage multi-task deep convolutional networks*

Sumber: [30]

Berdasarkan Gambar 2.2, arsitektur MTCNN terdiri dari tiga tahapan jaringan yang bekerja secara bertingkat (*cascaded*) untuk melakukan seleksi dan penyempurnaan deteksi wajah [30]. Setiap tahap bertujuan untuk mengurangi kandidat yang salah dan memperbaiki posisi kotak deteksi (*bounding box*) secara bertahap agar hasil akhir lebih akurat.

1. Tahap 1: Menggunakan *fully convolutional network* yang disebut *Proposal Network* (P-Net) untuk menghasilkan kandidat beserta *bounding box regression vectors*. *Regression vector* merupakan nilai koreksi numerik yang digunakan untuk menyesuaikan posisi dan ukuran *bounding box* agar lebih tepat terhadap lokasi wajah sebenarnya. Setelah itu, diterapkan metode *Non-Maximum Suppression* (NMS), yaitu teknik untuk menghapus *bounding box* yang saling tumpang tindih dan hanya mempertahankan kandidat dengan skor tertinggi.
2. Tahap 2: Seluruh kandidat dari P-Net diproses kembali oleh *Refine Network* (R-Net). Tahap ini berfungsi untuk menyaring lebih banyak kandidat palsu (*false positive*), melakukan kalibrasi ulang menggunakan *bounding box regression*, serta kembali menerapkan NMS untuk memastikan hanya kandidat terbaik yang dipertahankan.
3. Tahap 3: kandidat yang tersisa diproses oleh *Output Network* (O-Net). Memiliki mekanisme yang serupa dengan tahap sebelumnya, untuk mendeskripsikan wajah secara lebih rinci. Pada tahap ini, jaringan tidak hanya menghasilkan *bounding box*, tetapi juga memprediksi posisi lima titik fitur penting (*facial landmarks*) wajah yang berguna untuk proses penyalarsan (*alignment*) wajah sebelum tahap pengenalan.



Gambar 2.3. Alur proses deteksi wajah menggunakan MTCNN

Sumber: [30]

Gambar 2.3 menunjukkan alur kerja MTCNN secara keseluruhan. Sebelum masuk ke tahap P-Net, citra terlebih dahulu diubah ukurannya ke berbagai skala untuk membentuk *image pyramid*. Proses ini disebut *multi-scale detection*, yang bertujuan agar wajah dengan ukuran berbeda (besar maupun kecil) tetap dapat terdeteksi. Setiap skala citra kemudian diproses secara bertahap melalui P-Net, R-Net, dan O-Net hingga diperoleh hasil deteksi wajah yang paling akurat.

2.5 Contrast Limited Adaptive Histogram Equalization (CLAHE)

Contrast Limited Adaptive Histogram Equalization (CLAHE) adalah metode yang umum digunakan dalam peningkatan citra karena metodenya sederhana dan memiliki beban komputasi yang rendah [31]. Pada penelitian ini, CLAHE digunakan untuk meningkatkan kontras citra wajah. Penerapan *Contrast Limited Adaptive Histogram Equalization* (CLAHE) sebagai metode peningkatan kualitas citra mampu menonjolkan detail wajah, sehingga tetap terlihat jelas meskipun dalam kondisi pencahayaan terbatas (gambar skala abu-abu) [17]. Menonjolkan detail wajah, CLAHE dapat menggunakan mekanisme *clip limit* untuk membatasi histogram lokal, sehingga *noise* pada area seragam tidak diperkuat secara berlebihan dan detail penting pada wajah tetap dapat ditonjolkan tanpa menambah *noise* [32].

CLAHE beroperasi pada area kecil citra yang disebut *tile*. Pada setiap *tile*, tingkat kontras disesuaikan sehingga histogram pada area tersebut mengikuti bentuk yang telah ditetapkan. Selanjutnya, antar-*tile* digabungkan dengan *interpolasi bilinear*, yaitu metode *interpolasi* yang menghitung nilai piksel baru berdasarkan rata-rata tertimbang dari empat piksel tetangga terdekat secara dua arah (horizontal dan vertikal), sehingga transisi antar blok menjadi halus dan tidak menimbulkan batas yang terlihat jelas [33].

$$\beta = \frac{M}{N} \left(1 + \frac{\alpha}{100} (S_{\max} - 1) \right) \quad (2.3)$$

Keterangan variabel:

- (a) β : nilai batas (*clip limit*)
- (b) M : ukuran area
- (c) N : nilai level keabuan (256)
- (d) α : faktor klip, tambahan batas histogram dengan nilai 1 sampai 100
- (e) S_{\max} : kemiringan maksimum yang diperbolehkan

2.6 FaceNet

FaceNet adalah metode representasi wajah yang dirancang untuk menghasilkan *embedding* yang bersifat diskriminatif, sehingga dapat memudahkan proses pencocokan antar wajah. Dengan demikian, FaceNet memungkinkan identifikasi dan verifikasi wajah dilakukan dengan lebih tepat dan akurat [34]. Model ini mengambil citra wajah sebagai input dan mengubahnya menjadi sebuah vektor berisi 128/512 fitur yang merepresentasikan wajah tersebut, sehingga membuat proses perhitungan selanjutnya lebih efisien dibandingkan dengan menggunakan citra wajah yang utuh [35].

Dalam prosesnya, *embedding* yang dihasilkan umumnya dinormalisasi terlebih dahulu menggunakan norma L2 agar memiliki skala yang seragam. Normalisasi ini bertujuan untuk memastikan bahwa perbandingan antar vektor lebih stabil dan tidak dipengaruhi oleh besar kecilnya nilai absolut fitur, melainkan oleh arah vektor tersebut. Secara matematis, diberikan vektor *embedding* wajah $\mathbf{v} = [v_1, v_2, \dots, v_n]$, maka proses normalisasi dilakukan sebagai berikut.

$$|\mathbf{v}|_2 = \sqrt{\sum_{i=1}^n v_i^2 + \varepsilon} \quad (2.4)$$

$$\mathbf{v}_{\text{norm}} = \frac{\mathbf{v}}{|\mathbf{v}|_2} \quad (2.5)$$

Keterangan variabel:

- (a) \mathbf{v} : vektor *embedding* wajah hasil ekstraksi fitur
- (b) v_i : elemen ke- i dari vektor *embedding*
- (c) n : dimensi vektor *embedding*
- (d) $|\mathbf{v}|_2$: norma L2 dari vektor *embedding*
- (e) ε : nilai konstanta kecil untuk mencegah pembagian dengan nol, dengan $\varepsilon = 1 \times 10^{-12}$
- (f) \mathbf{v}_{norm} : vektor *embedding* hasil normalisasi

Adapun langkah-langkah pelatihan model FaceNet yaitu [36].

1. Langkah 1: Secara acak memilih sebuah citra sebagai *anchor image* (gambar acuan).
2. Langkah 2: Secara acak pilih sebuah gambar dari orang yang sama dan tandai sebagai *positive image*.
3. Langkah 3: Secara acak menentukan citra dari orang lain dan menandainya sebagai *negative image*.
4. Langkah 4: Mengatur parameter pada FaceNet supaya citra positif berada lebih dekat dengan citra *anchor* dibandingkan dengan citra negatif. Dengan kata lain, jarak *anchor-positive* lebih kecil daripada jarak *anchor-negative*.



Gambar 2.4. *Triplet loss*

Sumber: [34]

Sebagaimana ditunjukkan pada Gambar 2.4, FaceNet dilatih menggunakan pendekatan *triplet loss*. Metode ini melibatkan tiga sampel citra sekaligus dalam satu proses pelatihan, yaitu *anchor*, *positive*, dan *negative*. Citra *anchor* dan *positive* berasal dari individu yang sama, sedangkan citra *negative* berasal dari individu yang berbeda [37].

Langkah terakhir merupakan proses pengenalan wajah. Pada tahap ini, *embedding* wajah yang diperoleh dari citra latih dibandingkan dengan *embedding* citra uji menggunakan *classifier* [34]. *Classifier* adalah metode atau algoritma yang digunakan untuk menentukan kelas atau identitas suatu data berdasarkan fitur yang telah diekstraksi. Dengan langkah ini, hasil *embedding* FaceNet tidak hanya digunakan untuk klasifikasi, tetapi juga dapat diaplikasikan dalam verifikasi, *clustering*, maupun identifikasi wajah.

2.7 *Human Resource Information System (HRIS)*

Human Resource Information System (HRIS) adalah sistem berbasis teknologi informasi yang dikembangkan untuk menunjang pelaksanaan fungsi-fungsi sumber daya manusia secara optimal [38]. Dalam penerapannya, aplikasi ini mampu mempercepat proses serta memberikan dampak signifikan terhadap pengambilan keputusan yang berkaitan dengan karyawan [39]. HRIS merupakan sistem terintegrasi yang mengatur berbagai informasi terkait sumber daya manusia, mulai dari data karyawan, rekam jejak pekerjaan, hingga perencanaan pengembangan karier, sehingga mampu mempermudah kegiatan administratif perusahaan secara menyeluruh [6]. Lebih lanjut, keberadaan HRIS juga berperan dalam meningkatkan kinerja karyawan yang berdampak pada produktivitas serta pencapaian tujuan organisasi [7].