

## **BAB II**

### **TINJAUAN PUSTAKA DAN LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

Penelitian relevan yang menjadi tinjauan pustaka dalam penelitian ini adalah penelitian dari Luisa F. Polana, Satyajit Gupte (2019) [29], Ruining Feng (2020) [30], Zheng Wang, Hongyan Quan (2019) [31], Xuan Hau Nguyen, Thai Son Tran, Van Thinh Le, KimDuy Nguyen, Dinh-Tu Truong (2021) [17], penelitian Huijing Zhan, Boxin Shi, Jiawei Chen, Ling-Yu Duan, Alex C. Kot (2019) [32], penelitian Ulla Delfana Rosiani, Rosa Andrie Asmara, dan Nadhifatul Laeily (2019) [28], dan penelitian William Hutamaputra dan Fitri Utaminigrum (2021) [27].

Penelitian pertama oleh Luisa F. Polana, Satyajit Gupte (2019) yang berjudul “*Learning Fashion Compatibility Across Apparel Categories for Outfit Recommendation*”. Tujuan dari penelitian ini adalah untuk merekomendasikan *item* pakaian pelengkap yang cocok dengan kueri sebagai pembentuk pakaian yang populer. Metode yang digunakan *Siamese network* untuk ekstraksi fitur, dan CNN untuk mempelajari fitur warna. Hasil penelitian menunjukkan bahwa model dapat mempelajari warna dan hubungan gaya antara kategori pakaian sampai batas yang baik, serta pada setiap *item* memiliki kompatibilitas skor yang tinggi cenderung memiliki warna netral seperti putih dan hitam [29].

Persamaan penelitian ini dengan penelitian milik Luisa F. Polana, Satyajit Gupte (2019) adalah sama-sama memiliki tujuan untuk memberikan rekomendasi suatu pakaian. Perbedaan penelitian ini dengan penelitian milik Luisa F. Polana, Satyajit Gupte (2019) terletak pada teknik yang digunakan serta implementasi suatu model yang dihasilkan terhadap objek. Penelitian ini menggunakan teknik VGG16 untuk model rekomendasi pakaian dan metode autoencoder dan *Facial Landmark Points* yang melakukan penukaran wajah. Sedangkan penelitian milik Luisa F. Polana, Satyajit Gupte (2019) hanya mengategorikan setiap item pakaian yang tepat untuk menjadi rekomendasi dengan menggunakan algoritma CNN [29].

Penelitian kedua oleh Ruining Feng tahun 2020 yang berjudul “*To Become*

*Fashionable: A brief Review of Outfit Compatibility*” dengan tujuan untuk mencocokkan pakaian yang modis dan layak serta menganalisis pakaian yang populer sehingga dapat memberikan rekomendasi pakaian yang paling memuaskan bagi pelanggan. Penelitian ini menggunakan algoritma kompatibilitas, algoritma rekomendasi, *Neural Network*, *Graph Deep Learning*. Hasil penelitian berupa perbandingan dari beberapa model yang menggunakan 4 set kumpulan data yaitu data mode amazon, data polivora, data mode dalam, dan data IQON [30].

Persamaan penelitian ini dengan penelitian milik Ruining Feng tahun 2020 adalah sama-sama menggunakan tujuan untuk memberikan rekomendasi *fashion outfit*. Perbedaan penelitian ini dengan penelitian milik Ruining Feng (2019) adalah pada teknik yang digunakan, serta proses rekomendasi pakaian. Penelitian milik Ruining Feng menggunakan teknik *Neural Network* dan *Graph Deep Learning*, serta proses rekomendasi pakaian hanya mencocokkan pakaian yang sesuai ke konsumen tidak secara *realtime* sedangkan penelitian ini menggunakan VGG16 untuk model rekomendasi pakaian dan teknik autoencoder dan *Facial Landmark Points* untuk model penukar wajah konsumen sehingga dari hasil rekomendasi pakaian dapat dilakukan penukaran wajah secara *realtime* [30].

Penelitian ketiga oleh Zheng Wang, Hongyan Quan (2019) berjudul “*Fashion Outfit Composition Combining Sequential Learning and Deep Aesthetic Network*” yang bertujuan untuk menyusun pakaian yang tepat dengan menjamin bahwa pakaian yang dihasilkan sesuai dengan estetika seseorang. Penelitian ini menggunakan metode *Sequential Learning* dengan Bi-LSTM untuk mempelajari kompatibilitas mode serta menggunakan *Deep Aesthetic Networks* yang memanfaatkan jaringan saraf subjektif sehingga mampu mengekstraksi fitur estetika dari pakaian. Hasil penelitian ini model yang diusulkan mampu menyusun pakaian estetika dengan menunjukkan kompatibilitas pakaian yang diberikan [31].

Persamaan penelitian ini dengan penelitian milik Zheng Wang, Hongyan Quan(2019) adalah sama-sama menggunakan objek pakaian, serta persamaan pada tujuan yang dihasilkan adalah merekomendasikan sebuah pakaian kepada konsumen. Perbedaan penelitian ini dengan penelitian Zheng Wang, Hongyan Quan (2019) adalah pada metode yang digunakan serta implementasi suatu model.

Penelitian ini menggunakan metode VGG16 untuk melakukan tugas klasifikasi pakaian sehingga menghasilkan model rekomendasi pakaian. Kedua menggunakan teknik autoencoder dan *Facial Landmark Points* yang akan menukar wajah secara otomatis, sedangkan penelitian milik Zheng Wang, Hongyan Quan menggunakan metode *Sequential Learning Bi-LSTM* dan *Deep Aesthetic Networks* [31].

Penelitian keempat oleh Xuan Hau Nguyen, Thai Son Tran, Van Thinh Le, Kim Duy Nguyen, Dinh-Tu Truong (2021) dengan judul “*Learning spatio-temporal features to detect manipulated facial videos created by the Deepfake techniques*” [5] yang bertujuan untuk mendeteksi suatu *video* palsu atau kebohongan hasil dari teknologi *deep fake*. Teknik yang digunakan adalah autoencoder dengan metode CNN-3D. Hasil penelitian ini model mampu memberikan deteksi dengan *deep fake Face Forensic* dan kumpulan data VidTIMIT terutama dengan *video* berkualitas tinggi kinerja model mampu mengungguli kedua set data diatas 99% [17].

Persamaan penelitian ini dengan penelitian milik Xuan Hau Nguyen, Thai Son Tran, Van Thinh Le, Kim Duy Nguyen, Dinh-Tu Truong (2021) adalah sama-sama menggunakan metode CNN dengan teknik autoencoder. Sedangkan perbedaan terletak pada objek yang diteliti. Penelitian milik Xuan Hau Nguyen, Thai Son Tran, Van Thinh Le, Kim Duy Nguyen, Dinh-Tu Truong (2021) objek yang digunakan adalah suatu *video* yang telah dimanipulasi. Pada penelitian ini menggunakan objek wajah yang akan disintesis menggunakan teknik autoencoder atau dengan teknik *Facial Landmark Points* yang dilakukan proses implan wajah sumber dan wajah target antara wajah *realtime* dengan wajah yang ada di *video* [17].

Penelitian kelima oleh Huijing Zhan, Boxin Shi, Jiawei Chen, Ling- Yu Duan, Alex C. Kot (2019) dengan judul “*Fashion Recommendation On Street Images*” [13]. Tujuan penelitian ini adalah untuk memberikan rekomendasi *item fashion* dengan mempelajari hubungan kompatibilitas. Metode yang digunakan dalam penelitian ini adalah CNN yang melakukan pembelajaran fitur serta menggunakan Bi-LSTM untuk memodelkan hubungan kompatibilitas antara *item-item*. Hasil penelitian membandingkan kinerja metode Bi-LSTM dengan CSN berbasis pembelajaran matrik dan mendapatkan skor akurasi AUC *random negative* mencapai

95% dan *hard negative* mencapai 96% [32].

Persamaan penelitian ini dengan penelitian milik Huijing Zhan, Boxin Shi, Jiawei Chen, Ling-Yu Duan, Alex C. Kot (2019) adalah sama-sama memiliki tujuan untuk merekomendasikan pakaian. Sedangkan perbedaan pada cara kerja model CNN yang diterapkan. Penelitian milik Huijing Zhan, Boxin Shi, Jiawei Chen, Ling-Yu Duan, Alex C. Kot (2019) melakukan pengkategorian setiap *item* pakaian yang menjadi bahan rekomendasi untuk *user*, sedangkan penelitian ini melakukan pengujian model penukaran wajah dengan sebuah perangkat kamera yang akan mengambil objek wajah konsumen dan mengganti wajah model pakaian rekomendasi dengan wajah konsumen [32].

Penelitian keenam oleh Ulla Delfana Rosiani, Rosa Andrie Asmara, dan Nadhifatul Laeily yang berjudul “Penerapan *Facial Landmark Point* untuk Klasifikasi Jenis Kelamin Berdasarkan Citra Wajah”. Penelitian ini untuk mendeteksi jenis kelamin dengan menggunakan fitur geometri yang merupakan fitur pada *landmark* dan cocok untuk mengatasi kompleksitas. Metode yang digunakan untuk pengukuran nilai-nilai pada fitur dan metode *Support Vector Machine* untuk perhitungan klasifikasi. Hasil dari penelitian menggunakan *dataset* 60 data untuk *training* dan 20 untuk *testing* mendapatkan akurasi 70%, sedangkan menggunakan *dataset* 30 mendapatkan nilai akurasi 65% [28].

Persamaan penelitian Ulla Delfana Rosiani, Rosa Andrie Asmara, dan Nadhifatul Laeily yaitu pada metode yang digunakan adalah *landmark point* untuk pengukuran fitur geometri. Namun terdapat perbedaan pada metode klasifikasi yang digunakan, studi kasus yang pecahkan, serta penggunaan *landmark point*. Metode pada penelitian milik Ulla Delfana Rosiani, dkk menggunakan metode klasifikasi SVM sedangkan peneliti menggunakan CNN VGG16. Kemudian penelitian Ulla Delfana Rosiani menggunakan studi kasus untuk mendeteksi jenis kelamin sedangkan peneliti melakukan penukaran wajah untuk rekomendasi pakaian. Penerapan *Facial Landmark Point* pada penelitian Ulla Delfana Rosiani digunakan untuk pengukuran nilai-nilai fitur sedangkan peneliti menggunakan *Facial Landmark Point* untuk melakukan penukaran wajah [28].

Penelitian Ketujuh yang berjudul “Implementasi *Facial Landmark* dalam Pengenalan Wajah pada Sistem Pembayaran Elektronik” oleh Wiliam Hutamaputra dan Fitri Utaminingrum (2021). Penelitian ini bertujuan membuat autentikasi yang berbeda menggunakan sistem biometrik dengan wajah untuk transaksi elektronik. Penelitian ini menggunakan metode *Facial Landmark Points* untuk mendeteksi lokalisasi titik-titik yang menonjol pada wajah dan metode klasifikasi *K-Nearest Neighbor* untuk pengujian ketepatan pengenalan wajah. Hasil dari penelitian ini bahwa klasifikasi untuk pengujian ketepatan pengenalan wajah menggunakan  $K=5$  mendapatkan akurasi tertinggi pada setiap perubahan pada  $K$  nya yaitu 71,43% [27].

Persamaan penelitian ini dengan peneliti adalah pada metode yang digunakan yaitu *Facial Landmark Point* untuk sistem biometrik lokalisasi titik-titik pada wajah. Namun terdapat perbedaan pada metode klasifikasi yang digunakan, studi kasus, serta penerapan *Facial Landmark Point*. Metode klasifikasi yang digunakan oleh Willian Hutamaputra dan Fitri Utaminingrum adalah *K-Nearest Neighbor* sedangkan peneliti menggunakan metode CNN dengan *Transfer Learning* VGG16. Kemudian pada studi kasus, William Hutamaputra untuk pembuatan sistem pembayaran elektronik dengan pengenalan wajah. Sedangkan peneliti melakukan penukaran wajah untuk membuat rekomendasi pakaian yang cocok. Penelitian William Hutamaputra menggunakan *landmark point* untuk mendeteksi wajah dengan lokalisasi titik-titik yang menonjol sedangkan peneliti menggunakan *landmark point* untuk melakukan penukaran wajah [27].

Tabel 2.1 Tinjauan Pustaka

Judul, Peneliti	Metode & Objek	Kesimpulan	Saran
<i>Learning Fashion Compatibility Across Apparel Categories for Outfit Recommendation</i> oleh Luisa F. Polana, Satyajit Gupte (2019) [10].	Metode <i>Siamese Network</i> dan CNN. Objek <i>item fashion</i> pakaian.	Metode yang diusulkan menghasilkan rekomendasi untuk <i>item</i> pakaian pelengkap dengan diberikan <i>item</i> kueri.	Penelitian ini perlu ditambahkan kategori <i>outfit</i> bermotif dan polos sehingga rekomendasi yang diberikan oleh model lebih berkorelasi terhadap jenis pakaiannya.
<i>To Become Fashionable: A Brief Review of Outfit Compatibility</i> oleh Ruining Feng (2020) [11].	Algoritma kompatibilitas, algoritma rekomendasi, <i>neural network</i> , dan <i>Graph Deep Learning</i> . Objek <i>item fashion</i> pakaian.	Penelitian ini menyajikan berbagai konsep kompatibilitas pakaian serta membandingkan beberapa model dalam empat kumpulan data yang berbeda untuk menunjukkan model mana yang optimal.	Perlu diberikan suatu hasil dari model yang mampu menyajikan pakaian yang kompatibilitas pada jurnal penelitian.

Judul, Peneliti	Metode & Objek	Kesimpulan	Saran
<i>Fashion Outfit Composition Combining Sequential Learning and Deep Aesthetic Network</i> oleh Zheng Wang, Hongyan Quan (2019) [12].	Metode RNN dan <i>parallel</i> model Bi-LSTM. Objek <i>item fashion</i> pakaian.	Penelitian ini menunjukkan hasil pakaian estetik yang Memenuhi preferensi konsumen sehingga mampu mengevaluasi suatu koleksi <i>item</i> fashion dan dapat membentuk pakaian yang serasi dan estetik.	Penelitian ini perlu dikategorikan pakaian jenis bermotif atau polos sehingga Hasil rekomendasi lebih kompatibilitas walaupun tidak dipasangkan kepada konsumen.
<i>Learning Spatio-Temporal Features to Detect Manipulated Facial Videos Created by the Deepfake Techniques</i> oleh Xuan Hau Nguyen, Thai Son Tran, Van Thinh Le, Kim Duy Nguyen, Dinh-Tu Truong (2021).	Metode Autoencoder algoritma CNN-3D. Objek <i>video</i> yang telah dimanipulasi atau mengandung <i>deep fake</i> .	Mampu memberikan deteksi dengan <i>deep fake Face Forensic</i> dan kumpulan VidTIMIT dengan akurasi diatas 99%.	Penerapan model ditambahkan fitur yang mampu mendeteksi berbagai jenis peragaan wajah baik itu secara <i>face2face</i> maupun menggunakan alat <i>neural textures</i> .
<i>Fashion Recommendation on Street Images</i> oleh Huijing Zahn, Boxin Shi, Jiawei Chen, Ling-Yu Duan, Alex C. Kot	Metode CNN dengan Bi-LSTM. Objek <i>item fashion</i> pakaian serta sepatu dan tas.	Mampu mendapatkan nilai akurasi kinerja metode Bi-LSTM dengan <i>crucible swelling number</i> yaitu 95% untuk <i>random negative</i> ,	Perlu ditambahkan alur proses penelitian agar lebih terstruktur, serta dalam kategori belum membedakan mengenai jenis pakaian melainkan hanya warna pakaian saja sehingga model <i>Deep</i>

Judul, Peneliti	Metode & Objek	Kesimpulan	Saran
(2019) [13].		dan 96% untuk <i>hard negative</i>	<i>Learning</i> kurang efektif untuk memberikan rekomendasi kepada konsumen.
Penerapan <i>Facial Landmark Point</i> untuk Klasifikasi Jenis Kelamin berdasarkan Citra Wajah oleh Ulla Delfana Rosiani, Rosa Andrie Asmara, dan Nadhifatul Laeily(2019)	Metode <i>Facial Landmark Point</i> dan <i>Support Vector Machine</i> . Objek yang digunakan adalah wajah.	Metode <i>Facial Landmark Point</i> dan SVM dapat melakukan klasifikasi jenis kelamin berdasarkan citra wajah, walaupun akurasi yang dihasilkan masih belum baik yaitu 70% dengan 60 <i>dataset</i> dan 40% dengan 30 <i>dataset</i>	Mencoba menggunakan metode <i>Deep Learning</i> untuk melakukan klasifikasi jenis kelamin dan belum ada <i>scatter plot multidimensional</i> yang ditampilkan berdasarkan fitur dari data yang digunakan.
Implementasi <i>Facial Landmark</i> dalam Pengenalan Wajah pada Sistem Pembayaran Elektronik oleh William Hutamaputra dan Fitri Utamingrum (2021)	Metode <i>Facial Landmark Point</i> dan <i>K-Nearest Neighbor</i> . Objek yang digunakan adalah wajah.	Metode <i>facial landmark</i> dan diujikan dengan metode <i>K-Nearest Neighbor</i> menggunakan nilai $K=5$ mendapatkan nilai akurasi sebesar 71,43%.	Melakukan pendeteksian wajah dan pemberian titik dari sudut pandang yang berbeda, mencoba menggunakan metode klasifikasi <i>Deep Learning</i> seperti CNN, serta parameter dari sudut pandang wajah yang dikembangkan lebih banyak dan bervariasi untuk meningkatkan akurasi.

Berdasarkan tabel di atas terdapat penelitian yang serupa dengan penelitian ini. Oleh karena itu, meskipun terdapat kesamaan dengan penelitian lain, namun mengingat terdapat perbedaan pada teknik dan metode yang digunakan sehingga peneliti tertarik untuk melakukan penelitian tentang model rekomendasi pakaian



kepada konsumen dan model penukar wajah yang mendukung model rekomendasi agar lebih realistis seolah-olah konsumen mencoba pakaian tersebut padahal hanya dilakukan penukaran wajah antara wajah konsumen pada webcam dengan wajah video rekomendasi pakaian.

## 2.2 Landasan Teori

### 2.2.1 Pakaian

Pakaian adalah kebutuhan pokok manusia karena fungsinya sebagai penutup aurat manusia, pelindung tubuh manusia, petunjuk identitas manusia, perhiasan manusia, membantu manusia dalam melakukan suatu pekerjaan khusus, serta menghilangkan perbedaan antar manusia. Pakaian selain berfungsi untuk menutupi tubuh manusia, pakaian juga sebagai sebuah model atau aksesoris serta produk budaya dalam masyarakat tertentu [33].

### 2.2.2 Wajah

Wajah merupakan bagian depan dari kepala yang meliputi area dari dahi hingga dagu dan terdiri dari mata, hidung, mulut, pipi, serta dagu [34]. Wajah menjadi bagian dari tubuh yang paling sering terlihat dan digunakan untuk mengidentifikasi atau mengenali [35]. Selain itu, wajah juga dapat mengekspresikan berbagai emosi dan perasaan manusia seperti senang, sedih, marah, atau takut [36]. Wajah memiliki pola unik pada setiap individu, yang memungkinkan orang lain untuk mengenali seseorang dari wajahnya [37]. Oleh karena itu, wajah sering dianggap sebagai aspek penting dalam komunikasi antar manusia [38].

Penelitian ini melakukan rekonstruksi pada wajah sehingga menghasilkan wajah virtual. Wajah virtual pada representasi digital atau simulasi dari wajah manusia yang dibuat menggunakan teknologi komputer. Teknologi wajah virtual juga dapat digunakan untuk mengenali dan mereplikasi ekspresi wajah manusia secara *realtime*, sehingga memungkinkan interaksi manusia dengan karakter virtual yang lebih alami dan realistis [39].

### 2.2.3 Teknik Autoencoder

Teknik autoencoder digunakan untuk membuat *deep fake* [18]. Autoencoder

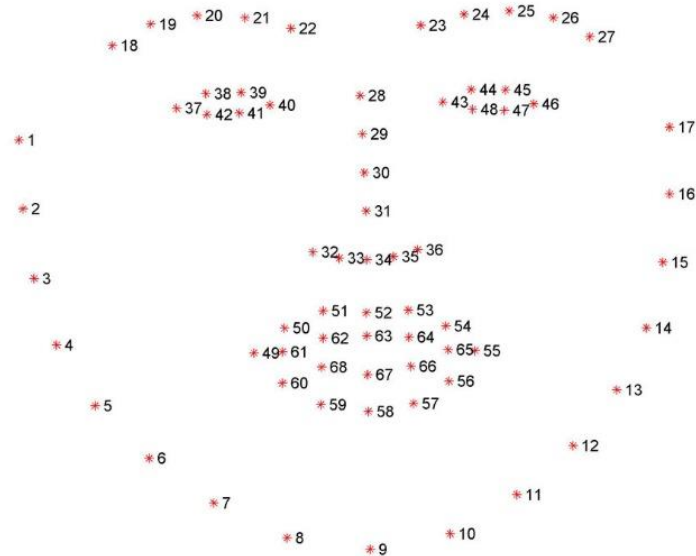
merupakan jaringan yang dilatih dengan metode *Supervised Learning*, tetapi membutuhkan kelas dari data sebagai target [40]. Autoencoder model *Neural Network* yang memiliki masukan dan keluaran yang sama dengan bobot yang sama antara encoder dan decodernya atau disebut dengan *tied weight* [41]. Kegunaan autoencoder adalah untuk mengurangi dimensi dari fitur yang didapat dari suatu Masukan atau bisa disebut dengan *dimensionality reduction* dan dimanfaatkan sebagai pemotong jaringan pada bagian encoder yang nantinya akan di *latent feature* [42]. Autoencoder juga dapat digunakan untuk menyelesaikan kasus derau atau disebut dengan *denoising* autoencoder [40].

#### 2.2.4 *Facial Landmark Points*

*Facial Landmark Points* adalah titik penting pada wajah manusia yang tujuannya untuk mengidentifikasi bagian wajah seperti mata, hidung, mulut, dan dahi [43]. *Landmarks points* beragam jenisnya, salah satu jenis *landmarks* yang digunakan dalam penelitian ini adalah *landmark 68 point* [44]. Beberapa titik *landmark* yang umum digunakan adalah

- a. Alis berada pada garis alis bagian atas dan bawah dan biasanya terdapat 5 titik pada setiap alis
- b. Mata berada pada bagian kelopak mata, bagian dalam mata, dan bagian bawah mata dan terdapat 6 sampai 8 titik.
- c. Hidung berada pada bagian atas hidung, tulang hidung, dan bagian bawah hidung serta biasanya terdapat 4 sampai 5 titik.
- d. Mulut berada pada bagian bibir atas dan bawah, sudut mulut, dan tengah mulut serta biasanya terdapat 12 sampai 14 titik.
- e. Pipi berada pada tulang pipi dan sampai wajah serta terdapat 3 sampai 4 titik [45].

Berikut ini gambar penempatan titik-titik *landmark 68 point* yang terdapat pada wajah



Gambar 2.1 *Facial Landmark Point* (sumber: shekharpandey, 2023)

Gambar 2.1 diatas adalah penempatan titik koordinat jenis *Facial Landmark Point* 68 pada wajah. Penempatan titik koordinat mampu melakukan ekstraksi fitur-fitur bagian wajah yang dibutuhkan bahkan semua bagian wajah [46].

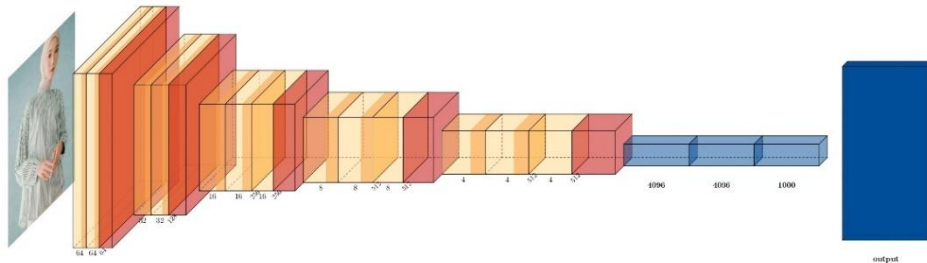
### 2.2.5 *Convolutional Neural Network*

*Convolutional Neural Network* (CNN) merupakan jaringan saraf yang mampu mengekstraksi fitur dari data dengan struktur konvolusi [47]. CNN memiliki kernel yang dapat merespon berbagai fitur seperti fitur aktivasi yang mensimulasikan sinyal listrik yang melebihi ambang batas akan dapat ditransmisikan ke *neuron* berikutnya. Keunggulan dari CNN yaitu memiliki koneksi lokal atau setiap *neuron* tidak lagi terhubung ke semua *neuron* pada lapisan sebelumnya sehingga mengurangi parameter dan mempercepat konvergensi. Keunggulan selanjutnya bahwa CNN dapat melakukan pembobotan yang sama serta dapat mengurangi dimensi *down sampling* sehingga dapat mengurangi jumlah data namun mempertahankan informasi yang berguna [48].

### 2.2.6 VGG16

VGG16 atau biasanya disingkat dengan VGG16 merupakan bagian dari *Convolutional Neural Network* (CNN) dengan model *Transfer Learning* atau model yang sebelumnya telah dilatih, digunakan kembali untuk menyelesaikan tugas baru [49]. VGG16 terdiri dari 16 lapisan konvolusi dimana 13 lapisan pertama adalah

lapisan konvolusi dan 3 lapisan terakhir adalah lapisan *fully connected* [50]. Urutan dari layer pada arsitektur VGG16 dapat dilihat pada Gambar 2.2. Penggunaan VGG16 dapat mempercepat proses pelatihan model, menghasilkan model yang baik walaupun dengan data yang sedikit, dan dapat menghindari *overfitting* pada data kecil [51]. VGG16 diimplementasikan untuk klasifikasi citra dan deteksi objek pada beberapa kumpulan data pembandingan termasuk kumpulan data imagenet yang sangat besar [52]. Arsitektur VGG16 ini terdiri dari beberapa *layer* sesuai Gambar 2.2 berikut ini.



Gambar 2.2 Arsitektur VGG16

Dimana pada Gambar 2.2 *block* berwarna kuning merupakan lapisan *convolutional layer*, blok berwarna *orange* merupakan *max pooling layer*, *block* berwarna biru muda merupakan lapisan *fully connected* dan lapisan yang berwarna biru tua merupakan *output layer*. Penjelasan terkait setiap *layer* dari arsitektur Gambar 2.2 terdiri dari beberapa lapisan dengan spesifikasi setiap layer yang dapat dilihat pada Tabel 2.2.

Tabel 2.2 Layer Arsitektur VGG16

Type Layer	Spesifikasi
<i>Input layer</i>	224x224 piksel
<i>Convolutional layer</i>	64 filter dan kernel 3x3
<i>Convolutional layer</i>	64 filter dan kernel 3x3
<i>Max pooling layer</i>	size 2x2 dan stride 2
<i>Convolutional layer</i>	128 filter dan kernel 3x3
<i>Convolutional layer</i>	128 filter dan kernel 3x3

<i>Max pooling layer</i>	<i>size 2x2 dan stride 2</i>
<i>Convolutional layer</i>	<i>256 filter dan kernel 3x3</i>
<i>Convolutional layer</i>	<i>256 filter dan kernel 3x3</i>
<i>Convolutional layer</i>	<i>256 filter dan kernel 3x3</i>
<i>Max pooling layer</i>	<i>size 2x2 dan stride 2</i>
<i>Convolutional layer</i>	<i>512 filter dan kernel 3x3</i>
<i>Convolutional layer</i>	<i>512 filter dan kernel 3x3</i>
<i>Convolutional layer</i>	<i>512 filter dan kernel 3x3</i>
<i>Max pooling layer</i>	<i>size 2x2 dan stride 2</i>
<i>Convolutional layer</i>	<i>512 filter dan kernel 3x3</i>
<i>Convolutional layer</i>	<i>512 filter dan kernel 3x3</i>
<i>Convolutional layer</i>	<i>512 filter dan kernel 3x3</i>
<i>Max pooling layer</i>	<i>size 2x2 dan stride 2</i>
<i>Fully Connected layer</i>	<i>4096 neuron</i>
<i>Fully Connected layer</i>	<i>4096 neuron</i>
<i>Fully Connected layer</i>	<i>1000 neuron</i>
<i>Output layer</i>	<i>1000 neuron dan aktivasi softmax</i>

### 2.2.7 Klasifikasi Gambar

Klasifikasi gambar adalah proses pengelompokkan gambar ke dalam kategori atau kelas tertentu berdasarkan fitur atau karakteristik tertentu yang terdapat pada gambar tersebut [53]. Teknik klasifikasi gambar pada tugas ini memanfaatkan jaringan saraf konvolusional atau CNN. Untuk proses pembuatan model klasifikasi maka diperlukan data gambar, kemudian dilakukan proses *training*, validasi, dan proses terakhir *testing* [54]. Dalam klasifikasi gambar terdapat data gambar dengan diketahui kelasnya untuk digunakan melatih model CNN sehingga model dapat mempelajari pola-pola atau fitur yang ada pada gambar [55]. Selanjutnya model yang telah dilatih digunakan untuk memprediksi kelas atau *label* gambar yang belum diketahui berdasarkan pola-pola atau fitur yang ditemukan dalam gambar tersebut [56].

### 2.2.8 Augmentasi Data

Augmentasi data adalah teknik untuk menambah variasi dan jumlah data pelatihan yang sudah ada [57]. Tujuannya untuk meningkatkan keberagaman data, mengurangi *overfitting*, meningkatkan nilai akurasi saat pelatihan model, dan mengatasi masalah data apabila tidak seimbang [58]. Augmentasi data memiliki berbagai macam teknik diantaranya pergeseran atau *shift*, rotasi, *zoom*, pemotongan, pengubahan kecerahan, distorsi, *noise*, *blurring*, *color channel shifting* atau mengubah warna menjadi channel tertentu pada gambar [59]. Teknik-teknik augmentasi dapat digunakan secara bersamaan sehingga menghasilkan variasi gambar yang lebih banyak [60].

### 2.2.9 HAAR *Face Frontal*

Haar merupakan suatu metode deteksi objek berdasarkan *cascade classifier*. Haar *cascade* digunakan untuk mendeteksi objek gambar yang bersifat digital [61]. Haar *cascade* merupakan sebuah persegi *features* yang dapat memberikan indikasi spesifik terhadap objek sehingga dapat dikenali berdasarkan nilai sederhana dari fitur. Kelebihan dari metode Haar *Cascade* adalah memiliki proses komputasi yang sangat cepat [62]. Haar memiliki bermacam-macam jenis, salah satunya haar *face frontal* [63]. Haar *face frontal* akan mendeteksi atau mengindikasikan wajah seseorang dan akan memberikan *rectangular* pada mata hidung dan mulut [64].

### 2.2.10 *Loss*

*Loss* merupakan suatu matrik yang digunakan untuk mengevaluasi seberapa akurat model dalam memprediksi *output* atau target yang diharapkan [65]. Fungsi *loss* akan memperhitungkan selisih antara *output* prediksi model dengan target yang diharapkan, dan kemudian menghitung besarnya kerugian tersebut [66]. Semakin kecil nilai fungsi *loss* yang dihasilkan, maka semakin baik performa model dalam memprediksi target yang diharapkan [67].

Fungsi *loss* sering digunakan sebagai fungsi objektif dalam proses optimasi model, dimana tujuannya adalah untuk menemukan parameter-parameter model yang dapat meminimalkan nilai fungsi *loss*. Dalam optimisasi model, algoritma akan mencoba untuk menemukan kombinasi parameter yang dapat meminimalkan nilai

fungsi *loss* pada data latih, sehingga model dapat melakukan prediksi dengan akurasi yang lebih tinggi pada data yang belum dilihat sebelumnya [68].

#### 2.2.11 Bias

Bias merupakan kesalahan sistematis yang terjadi pada model saat mencoba mempelajari hubungan antara variabel *input* dan *output* [69]. Pengukuran bias dapat dilakukan dengan metode *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), atau *Root Mean Squared Error* (RMSE). Jika semakin tinggi nilai bias, maka kinerja model buruk dalam melakukan prediksi dan semakin besar kesalahan pada data pelatihan [70].

#### 2.2.12 Multi-Class Confusion Matrix

*Confusion matrix* adalah suatu tabel yang digunakan untuk mengevaluasi performa dari suatu model klasifikasi dengan membandingkan nilai prediksi model dengan nilai sebenarnya dari data yang telah diketahui [71]. Penggunaan dari *confusion matrix* yaitu untuk menunjukkan jumlah hasil prediksi yang benar dan yang salah dari setiap kelas target, serta membantu untuk mengukur kinerja model dalam melakukan klasifikasi [72]. Terdapat dua macam *confusion matrix* yaitu *binary confusion matrix* dan *multi-class confusion matrix*. *Confusion Matrix* 2 dimensi memiliki persamaan akurasi berikut ini.

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.2)$$

Keterangan:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Berdasarkan formula diatas, *true positive* artinya bahwa jumlah data yang sebenarnya dengan prediksi sama, *true negative* merupakan jumlah data yang sebenarnya dan prediksinya adalah *negative*, *false positive* merupakan jumlah data yang sebenarnya *negative*, namun setelah diprediksi hasilnya *positive*, sedangkan *false negative* terjadi ketika jumlah data yang sebenarnya adalah *positive*, tetapi diprediksi hasilnya *negative*.

Tahap evaluasi penelitian ini *confusion matrix* yang digunakan adalah multi-class confusion matrix. *Multi-class confusion matrix* digunakan terhadap kasus yang memilih lebih dari dua kelas target yang harus diprediksi. *Confusion matrix* ini memiliki bentuk matriks dengan baris dan kolom yang mempresentasikan kelas target, dan setiap elemen matriks merepresentasikan jumlah data yang terklasifikasi dengan benar atau salah [73].