

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1. Tinjauan Pustaka

Pembelajaran mesin menuai popularitas pada beberapa tahun belakangan ini. Salah satu bagian dari pembelajaran mesin yang banyak digunakan adalah deep learning, yang layer pemrosesan berlapis untuk mempelajari representasi data dengan abstraksi multilevel (LeCun dkk., 2015). Deep learning populer digunakan untuk permasalahan pengenalan pola dan klasifikasi, antara lain pengenalan wajah. Teknologi deep learning telah memodelkan dengan sukses fitur wajah yang kompleks dengan disediakan data citra yang berkualitas (Arachchilage & Izquierdo, 2020).

Deep convolutional neural networks (DCNNs) merupakan pendekatan holistik yang pada masa ini memungkinkan terjadinya lompatan besar pada bidang pengenalan wajah. Pada tahun 2014, Facebook melaporkan sistem pengenalan wajah yang diberi nama DeepFace (Taigman, dkk, 2014) yang mampu mencapai performa mendekati manusia pada LFW benchmark (Huang, dkk, 2008).

Tingkat akurasi tersebut dengan cepat mampu dilampaui oleh sistem seperti DeepId3 (Sun, dkk, 2015) dan FaceNet (Schroff, dkk, 2015). Kemajuan yang substansial dari teknologi pengenalan wajah merupakan hasil dari penelitian dan pengembangan dalam arsitektur deep learning selama bertahun-tahun. Dimulai dari LeNet pada 1989 (Lecun, dkk, 1998), DCNNs telah berevolusi menjadi jaringan yang luar biasa terutama karena dorongan untuk menjawab tantangan klasifikasi, seperti The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky, dkk, 2015). AlexNet (Krizhefsky, dkk, 2012), VGGNet (Simonyan & Zisserman, 2015), dan GoogleNet (Szegedy, dkk, 2015) dikenal sebagai tiga jaringan paling berpengaruh pada ILSVRC.

Hasil penelitian dari Mark Egan, dkk. (2021) menunjukkan bahwa kampanye grafis di ruang publik mengenai pentingnya penggunaan masker wajah untuk mencegah penularan Covid-19 terbukti efektif meningkatkan kesadaran publik. Joshi, dkk (2020) membangun deep learning framework untuk mengidentifikasi penggunaan masker dengan menggunakan Multi-Task Cascaded Convolutional Neural Network (MTCNN) model (Zhang, dkk, 2014) dengan tingkat akurasi 81,74%. Deep learning bekerja dengan baik dengan menggunakan data latih yang berjumlah besar. Namun ada kalanya dataset dalam jumlah besar tidak tersedia. Ketidaktersediaan data ini menyebabkan sulitnya membentuk model yang akurat

sebagaimana dihasilkan oleh dataset yang besar. Wang, dkk (2020) mencoba berperan dalam pengumpulan dataset wajah bermasker untuk menjawab tantangan ini. Hanya saja, jika dibandingkan dengan proses manusia belajar, pada umumnya manusia tidak belajar suatu hal benar-benar dari nol. Sebagai contoh, apabila seseorang telah mempelajari cara naik sepeda, maka pada saat orang tersebut akan belajar naik motor, ia tidak benar-benar belajar dari awal lagi, melainkan telah memiliki pengetahuan sebelumnya dari belajar naik sepeda. Hal inilah yang menjadi kunci pada teknik transfer learning. Transfer learning merujuk pada sebuah cara training dengan menggunakan model yang sudah dilatih sebelumnya dan selanjutnya menggunakan model tersebut untuk sebuah dataset baru (Goodfellow, dkk, 2016). Penelitian ini bertujuan untuk mendeteksi masker dengan menggunakan model transfer learning yang populer digunakan, yaitu MobileNetV2.

Peneliti	Topik	Metode	Perbedaan
Nirna Tiskadewi (2017)	Identifikasi Citra Iris Mata	<i>KNN (K-Nearest Neighbor)</i>	Metode, dataset, pemrograman
Imam Taufiq (2018)	Deteksi Tanda Nomor Kendaraan Bermotor	Convolutional Neural Network Dengan Python Dan Tensorflow	Metode, dataset, pemrograman
Sun, dkk, (2015) dan Schroff, dkk, (2015)	Teknologi pengenalan wajah (DeepId3 dan FaceNet)	Deep Learning	Metode, dataset, pemrograman
Joshi, dkk (2020)	Mengidentifikasi penggunaan masker	Multi-Task Cascaded Convolutional Neural Network (MTCNN)	Metode, dataset, pemrograman
Oumina, A., Makhfi, N. E., dan Hamdi, M., (2020)	Mendeteksi penggunaan masker wajah	Transfer Learning (MobileNetV2, VGG19, Xception), jumlah dataset 1.376 gambar dengan tingkat akurasi 97,1 %	Dataset : jumlah gambar 7.553. Tingkat akurasi: 99%

Tabel 2.1. Perbandingan dengan penelitian yang sudah ada

## 2.2. Dasar Teori

### 2.2.1. Pembelajaran Mesin

Pembelajaran mesin atau machine learning merupakan suatu metode dari artificial intelligence yang membangun suatu model statistika berdasarkan pola data yang diberikan

untuk mendapatkan keputusan dalam memprediksi dan mengklasifikasi suatu data baru yang datang di masa depan menurut Samuel (1959).

Algoritma pembelajaran mesin membangun model matematika berdasarkan data sampel, yang dikenal sebagai "data pelatihan", untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk melakukan tugas. Pembelajaran mesin terkait erat dengan statistik komputasi, yang berfokus pada membuat prediksi menggunakan komputer. Beberapa pendekatan pengenalan pola yang didasarkan pada penggunaan machine learning yaitu klustering, klasifikasi, regresi, rekognisi, dan lain sebagainya. Klasifikasi merupakan proses mengidentifikasi suatu data ke dalam suatu kelas dalam objek yang berbeda, sedangkan rekognisi sama seperti klasifikasi, hanya memiliki perbedaan dalam objek yang sama. Dimisalkan jika klasifikasi dapat mengidentifikasi citra angka ke bilangan angka, sedangkan rekognisi dapat mengenali angka satu dari berbagai varian citra angka satu.

### **2.2.2. Deep Learning**

Deep learning merupakan bagian dari metode pembelajaran mesin yang menggunakan jaringan syaraf tiruan. Deep learning merupakan model komputasional yang terdiri dari layer pemrosesan berlapis untuk mempelajari representasi data dengan abstraksi multilevel (LeCun dkk., 2015). Deep learning berfokus pada penggunaan model jaringan syaraf tiruan besar dan cocok digunakan dalam konteks data yang kompleks dan tersedia dataset yang besar. Pada saat ini hampir semua perusahaan IT besar menggunakan deep learning, misalnya Facebook menggunakan deep learning untuk menganalisis teks dalam percakapan online, Google, Baidu, dan Microsoft menggunakan deep learning untuk pencarian image dan translasi mesin, deep learning pada saat ini juga merupakan teknologi yang standar digunakan untuk speech recognition pada smartphone, serta pengenalan wajah pada kamera (Kelleher, 2019). Komponen kunci berkaitan dengan deep learning adalah representasi data hirarkis yang berlapis-lapis (multilayered hierarchical data representation), yang biasanya berbentuk jaringan syaraf tiruan dengan lebih dari dua lapisan atau layer (Druzhkov & Kustikova, 2016).

Istilah multilayered hierarchical data representation merujuk pada representasi data yang dikembangkan dengan menggunakan banyak lapisan. Sebagai contoh kita bisa membayangkan dimana lapisan pertama mempelajari garis tepi atau warna yang ditemukan pada gambar, lapisan kedua mempelajari tentang bentuk atau sudut, dan lapisan tinggi lebih banyak mempelajari konsep yang lebih kompleks seperti bentuk, ukuran, dan bagian mana yang terhubung satu sama lain (McCoppin & Rizki, 2014). Terdapat tiga jenis lapisan dalam

deep learning yakni lapisan pertama merupakan input layer, lapisan terakhir merupakan output layer, dan lapisan-lapisan di tengahnya disebut dengan hidden layer. Masing-masing layer menerima input dari layer sebelumnya.

Ekstraksi fitur tradisional membutuhkan banyak pengetahuan data untuk mengetahui jenis fitur apa yang harus digunakan. Bahkan jika kita tahu fitur seperti apa yang harus diekstraksi, ada banyak parameter yang harus disesuaikan untuk mengekstrak fitur tersebut. Sedangkan deep learning berusaha untuk menghindari permasalahan tersebut dengan menentukan fitur secara langsung dari data. Proses pelatihan seperti ini memerlukan jumlah data yang besar dan memakan waktu yang lebih banyak. Peneliti menghindari masalah data yang besar dengan membuat data buatan augmentasi data. Sedangkan untuk mengatasi permasalahan waktu pemrosesan digunakanlah perangkat keras yang efisien dalam hal komputasi seperti penggunaan graphics processing units (GPU) ataupun superkomputer.

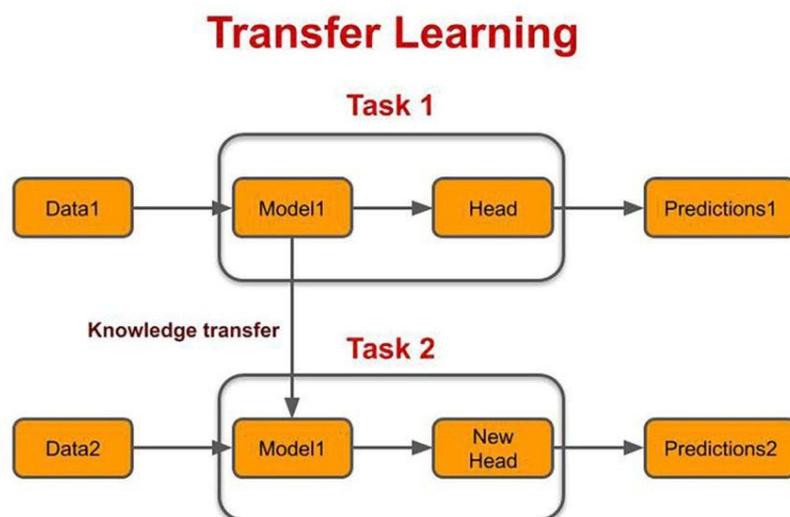
### **2.2.3. Transfer Learning**

Deep learning muncul untuk menjawab tantangan dari data yang berjumlah besar. Pada era digital seperti saat ini, data merupakan salah satu aset yang berharga. Namun ada kalanya dataset dalam jumlah besar tidak tersedia. Ketidakterersediaan data ini menyebabkan sulitnya membentuk model yang akurat sebagaimana dihasilkan oleh dataset yang besar. Di sisi lain, jika dibandingkan dengan proses manusia belajar, pada umumnya manusia tidak belajar suatu hal benar-benar dari nol. Sebagai contoh, apabila seseorang telah mempelajari cara naik sepeda, maka pada saat orang tersebut akan belajar naik motor, ia tidak benar-benar belajar dari awal lagi, melainkan telah memiliki pengetahuan sebelumnya dari belajar naik sepeda. Hal inilah yang menyebabkan para ilmuwan memprakarsai teknik transfer learning. Transfer learning merujuk pada sebuah cara training dengan menggunakan model yang sudah dilatih sebelumnya dan selanjutnya menggunakan model tersebut untuk sebuah dataset baru (Goodfellow, dkk, 2016). Model ini disebut juga sebagai pre-trained model karena telah dilatih dengan dataset umum sebelumnya.

Suatu pre-trained model adalah model yang sebelumnya telah dilatih pada suatu dataset dan telah memiliki bobot dan bias yang merepresentasikan fitur dari dataset yang digunakan untuk melatihnya. Fitur yang telah dipelajari tersebut dapat ditransfer ke data yang berbeda. Misalnya, model yang telah dilatih pada dataset yang besar dari gambar burung akan memiliki fitur yang telah dipelajari berupa sudut atau garis horizontal yang dapat ditransfer pada dataset yang baru.

Pada domain pengenalan wajah, dataset dapat menjadi faktor penentu pada performa model dikarenakan terbatasnya kemampuan manusia untuk memberi label secara manual pada data training (Oumina, dkk, 2020). Selain data dari domain target, data yang berhubungan dari domain lain pun dapat pula dimasukkan untuk memperluas ketersediaan dari pengetahuan sebelumnya terhadap target data yang dituju (Shao, dkk, 2015), sehingga pada penelitian ini digunakan deep learning model berupa *pre-trained model* MobileNetV2. Transfer learning mampu memberi akurasi yang baik dan mengatasi problem terbatasnya dataset serta permasalahan cross-domain dengan mengekstrak informasi berguna dari data yang telah ada dan ditransfer untuk digunakan pada data tujuan.

Transfer learning dari pre-trained model bekerja sebagaimana ditunjukkan pada Gambar 2.1. Task 1 pada Gambar 2.1 merupakan pre-trained model yang telah dilakukan terhadap ribuan dataset dari berbagai domain, termasuk di dalamnya adalah untuk image wajah. Kemudian pre-trained model tersebut menjadi knowledge yang ditransfer untuk Task 2, dalam hal ini adalah model untuk wajah dengan dan tanpa masker. Knowledge yang ditransfer tersebut, ditambah dataset baru yang spesifik untuk Task 2 (wajah dengan dan tanpa masker) akan membentuk sebuah model baru yang akan menjadi hasil pada penelitian Tugas Akhir ini.

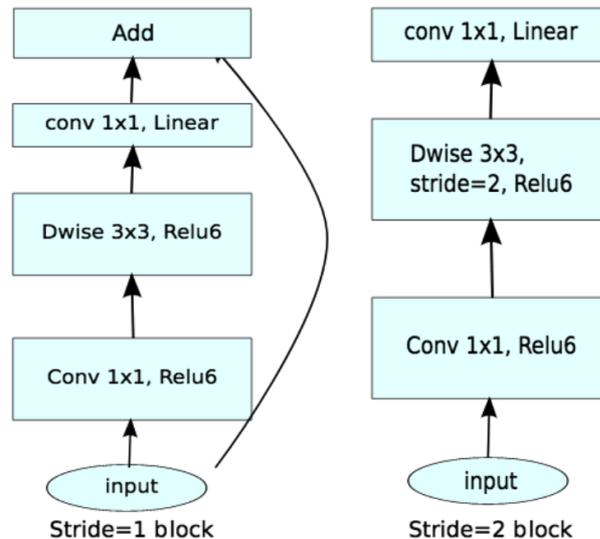


Gambar 2.1. Proses Transfer Learning

Untuk mendapatkan akurasi yang baik, transfer learning dari pre-trained model ke model tujuan perlu untuk dilakukan fine-tuning terhadap beberapa parameter. Parameter yang dapat dilakukan *fine-tuning* misalnya *learning rate*, *batch size*, dan *number of epochs* (Venu, 2020).

#### 2.2.4. MobileNetV2

MobileNetV2 merupakan model pengembangan dari MobileNetV1. Pada percobaan klasifikasi gambar menggunakan ImageNet, MobileNetV2 menunjukkan hasil akurasi yang lebih baik dibandingkan MobileNetV1 dengan jumlah parameter yang lebih sedikit (Sandler dkk, 2018). Pada MobileNetV2, terdapat dua jenis blok seperti yang tercantum pada gambar 2.4 di bawah ini yakni blok residual dengan stride 1 dan dengan stride 2, kedua blok tersebut disusun membentuk arsitektur MobileNetV2, seperti tampak pada Gambar 2.1.



Gambar 2.2. Arsitektur Dasar MobileNetV2

Sesuai namanya, MobileNetV2 merupakan arsitektur jaringan syaraf tiruan yang ditujukan agar memiliki performa yang baik pada perangkat mobile. MobileNetV2 didasarkan pada struktur *inverted residual* dimana koneksi residual berada antara bottleneck layers. Layer intermediate expansion menggunakan lightweight depthwise convolutions untuk memfilter fitur sebagai sumber dari non-linearitas. Secara keseluruhan, arsitektur MobileNetV2 terdiri dari initial fully convolution layer dengan 32 filter, diikuti dengan 19 residual bottleneck layer (Sandler, dkk 2018).

#### 2.2.5. Data Image Wajah

Wajah menyampaikan banyak fitur yang diskriminatif yang cukup kaya untuk menentukan identitas seseorang (Burrows dan Cohn, 2015). Fitur-fitur ini dapat diekstraksi dalam berbagai cara dan manipulasi yang non-intrusif. Oleh karena itu, pengenalan wajah otomatis dapat dimanfaatkan dalam banyak hal praktikal (Chellappa, dkk, 1995). Antara lain, penggunaannya yang memiliki performa luar biasa dalam aplikasi keamanan seperti

pengawasan cerdas, aplikasi otentikasi pengguna seperti verifikasi wisatawan di titik lintas batas, dan beragam aplikasi seluler dan media sosial lainnya (Arachchilage dan Izquierdo, 2020). Pemodelan dan pemetaan wajah yang mendasarinya sebenarnya rumit karena keragaman fitur wajah. Kompleksitas seperti itu semakin bertambah oleh variasi lain, seperti emosi, iluminasi, make up, dan penginderaan berkualitas rendah (Jia dan Gong, 2015).

Pada era pembelajaran mesin dan deep learning seperti sekarang, algoritma pendeteksi wajah telah mampu memberikan akurasi yang baik, jika diberikan dataset yang memadai. Pengumpulan dataset dan *sharing* dataset telah menjadi langkah yang dilakukan oleh peneliti di seluruh dunia secara kolektif. Salah satu repositori dan tempat sharing dataset yang banyak digunakan oleh para peneliti adalah kaggle (<https://www.kaggle.com>). Terdapat berbagai dataset wajah yang dikumpulkan dan dibagi oleh peneliti sebelumnya, dengan tujuan untuk memperkaya dataset kolektif dan untuk keberlanjutan penelitian di bidang tersebut.

#### **2.2.6. Preprocessing Image**

Dataset yang diunduh dari situs sharing dataset kaggle tersebut belum tentu memiliki kriteria yang sama dengan yang diinginkan. Oleh karena itu, perlu dilakukan pre-processing terlebih dahulu. Beberapa hal yang dilakukan pada tahap ini antara lain, konversi mode warna, perubahan ukuran image, dan augmentasi.

Perubahan ukuran image perlu dilakukan mengingat MobileNetV2 memiliki input size yang berbeda dengan data yang tersedia pada Kaggle. Sehingga image pada dataset perlu disesuaikan sebelum dapat diproses pada jaringan MobileNetV2.

Augmentasi adalah proses untuk meningkatkan variabilitas pada dataset yang bertujuan untuk menghindari overfitting. Overfitting adalah keadaan dimana model dapat bekerja dengan baik pada data training, sementara apabila model tersebut harus memprediksi data yang tidak ada pada data training (data uji), maka nilai akurasi akan berkurang secara signifikan. Data training akan dikenakan operasi seperti rotasi, resizing, shifting, dan lain-lain (Siegfried, 2021).

#### **2.2.7. Akurasi Pengujian**

Salah satu parameter untuk mengukur kinerja suatu model deep learning adalah akurasi. Akurasi menunjukkan jumlah prediksi yang benar dibandingkan dengan seluruh prediksi yang dilakukan. Adapun metode perhitungan akurasi yang akan dilakukan yaitu

dengan metode *Confusion Matrix* rumus Akurasi (*Accuracy*). Akurasi dapat dihitung dengan menggunakan Persamaan 1.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad \text{Persamaan 1}$$

Dimana:

TP = True Positive, yaitu prediksi bermasker terhadap image bermasker

TN = True Negative, yaitu prediksi tidak bermasker terhadap image tidak bermasker

FP = False Positive, yaitu prediksi bermasker terhadap image tidak bermasker

FN = False Negative, yaitu prediksi tidak bermasker terhadap image bermasker.

### 2.2.8. Presisi

Presisi merupakan salah satu pengukuran evaluasi kinerja model machine learning. Presisi merupakan rasio dari prediksi yang benar dari observasi positif (True Positive) terhadap seluruh prediksi positif yang dilakukan (True Positif maupun False Positive). Misalnya, untuk kelas bermasker, maka presisi dihitung dari rasio prediksi bermasker yang benar dibanding dengan seluruh prediksi bermasker. Sedangkan untuk kelas tanpa masker, maka presisi dihitung dari rasio prediksi tanpa masker yang benar dibanding dengan seluruh prediksi tanpa masker. Dalam bahasa yang lebih sederhana, presisi menjawab pertanyaan: “dari semua prediksi bahwa gambar ini memakai masker, berapakah yang memang benar bermasker?”

Presisi dapat dihitung dengan menggunakan Persamaan 2.

$$Precision = \frac{TP}{(TP+FP)} \quad \text{Persamaan 2}$$

Dimana:

TP = True Positive

FP = False Positive

### 2.2.9. Recall

Recall juga merupakan salah satu pengukuran evaluasi kinerja model machine learning. Recall disebut juga sebagai sensitivity, yaitu rasio dari prediksi positif yang benar (True positive) terhadap semua data pada kelas aktual (True Positif maupun False Positive).

Misalnya, untuk kelas bermasker, maka recall dihitung dari rasio prediksi bermasker yang benar dibanding dengan seluruh anggota kelas bermasker baik itu yang diprediksi bermasker maupun tidak bermasker. Sedangkan untuk kelas tanpa masker, maka recall dihitung dari rasio prediksi tanpa masker yang benar dibanding dengan seluruh anggota kelas tanpa masker baik itu yang diprediksi bermasker maupun tidak bermasker. Dalam bahasa yang lebih sederhana, recall menjawab pertanyaan: “dari semua gambar yang bermasker, berapa yang benar diprediksi bermasker?”

Recall dapat dihitung dengan menggunakan Persamaan 3.

$$Recall = \frac{TP}{(TP+FN)} \quad \text{Persamaan 3}$$

Dimana:

TP = True Positive

FN = False Negative

#### 2.2.10. F1 Score

F1 Score merupakan rata-rata terbobot dari Presisi dan Recall. Sehingga, nilai F1 score memperhitungkan baik false positive dan false negative. F1 score merupakan metode evaluasi model yang dapat lebih berguna dibandingkan akurasi apabila memiliki kelas dengan jumlah data yang tidak seimbang.

F1 score dapat dihitung dengan menggunakan Persamaan 4.

$$F1\ Score = \frac{2*(Recall*Precision)}{(Recall+Precision)} \quad \text{Persamaan 4}$$