

BAB 2

LANDASAN TEORI

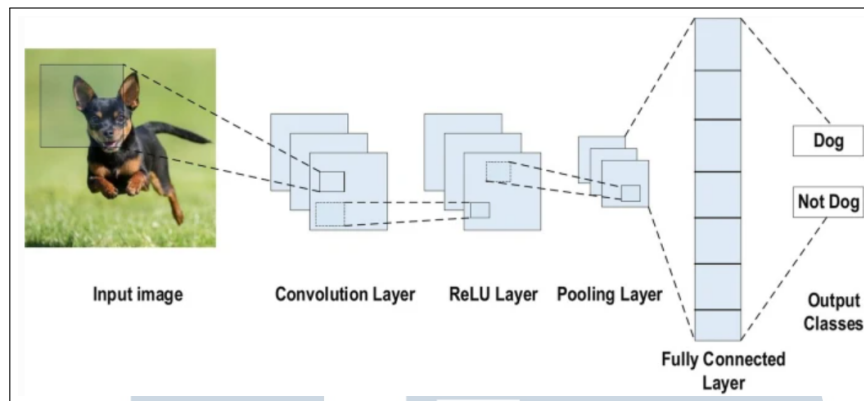
2.1 Deepfake

Deepfake adalah teknik memanipulasi citra manusia dengan menggunakan bantuan teknologi *Artificial Intelligence* (AI). Teknologi ini menggabungkan gambar atau video ke dalam sumber lain dengan menggunakan metode *machine learning* yang dikenal sebagai *Generative Adversarial Network* (GAN)[8]. Teknologi *deepfake* mampu memodifikasi ekspresi, identitas, atau karakteristik suatu subjek dengan mengubah wajah dan ekspresi seseorang. Proses ini memanfaatkan kecerdasan buatan serta teknik pemetaan wajah untuk menghasilkan gambar dan video yang tampak asli[9].

Seiring dengan kemajuan teknologi, penggunaan *deepfake* semakin meluas dan menjadi perhatian di berbagai bidang, terutama dalam aspek etika dan keamanan digital. Teknologi ini telah digunakan untuk membuat video palsu tokoh publik yang menyampaikan pernyataan yang sebenarnya tidak pernah diucapkan, sehingga menimbulkan potensi penyalahgunaan yang sangat besar. Selain itu, *deepfake* juga mulai digunakan dalam penipuan digital, penyebaran berita palsu, dan bahkan kejahatan siber lainnya. Hal ini menunjukkan bahwa meskipun *deepfake* memiliki potensi inovatif, tantangan terkait integritas informasi dan privasi individu menjadi isu penting yang perlu ditangani dengan serius[10].

2.2 Convolutional Neural Network

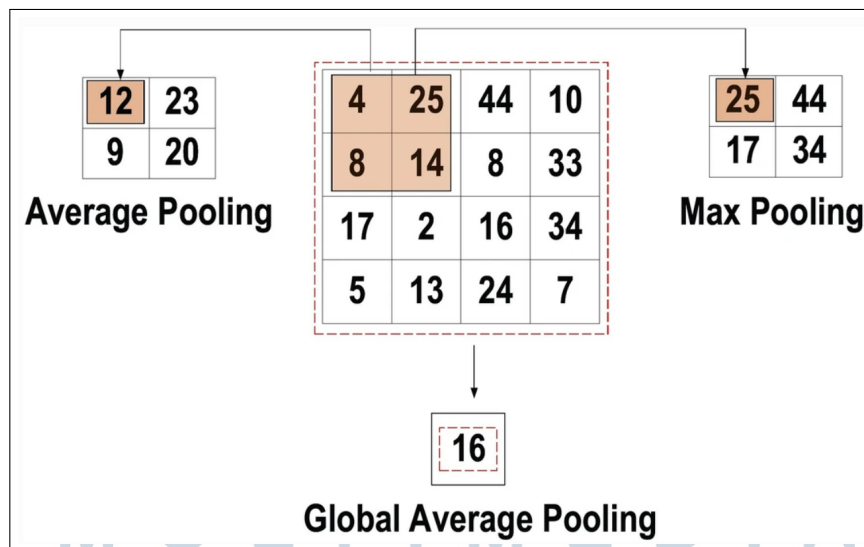
CNN adalah algoritma yang paling terkenal dalam bidang *Deep Learning* dan banyak digunakan dalam berbagai bidang seperti *computer vision*, pengolahan visual, dan pengenalan wajah. Manfaat utama *CNN* termasuk kemampuannya dalam mengidentifikasi fitur-fitur relevan secara otomatis dan memanfaatkan struktur data 2D seperti gambar. *CNN* memiliki struktur yang terinspirasi oleh otak manusia dan hewan, khususnya sel korteks visual. *CNN* yang umum digunakan menyerupai *multi-layer-perceptron* yang memiliki beberapa lapisan *convolutional layer*, *pooling layer*, dan *fully connected layer*. Arsitektur *CNN* dapat dilihat pada Gambar 2.1 [11].



Gambar 2.1. Arsitektur Convolutional Neural Network

Sumber: [11]

Convolution layer merupakan layer pertama setelah input gambar. Lapisan ini menggunakan *filter* untuk mengekstraksi fitur gambar, terdapat beberapa parameter yang diubah guna memodifikasi ciri-ciri lapisan yaitu ukuran *filter*, *stride*, dan *padding*. Fungsi aktivasi *ReLU* (*Rectified Linear Unit*) digunakan untuk memperkenalkan sifat non-linier pada model dan meningkatkan kemampuannya dalam merepresentasikan data. Fungsi ini didefinisikan sebagai $f(x) = \max(0, x)$, di mana output neuron akan bernilai nol jika input yang diterima bernilai negatif[12].

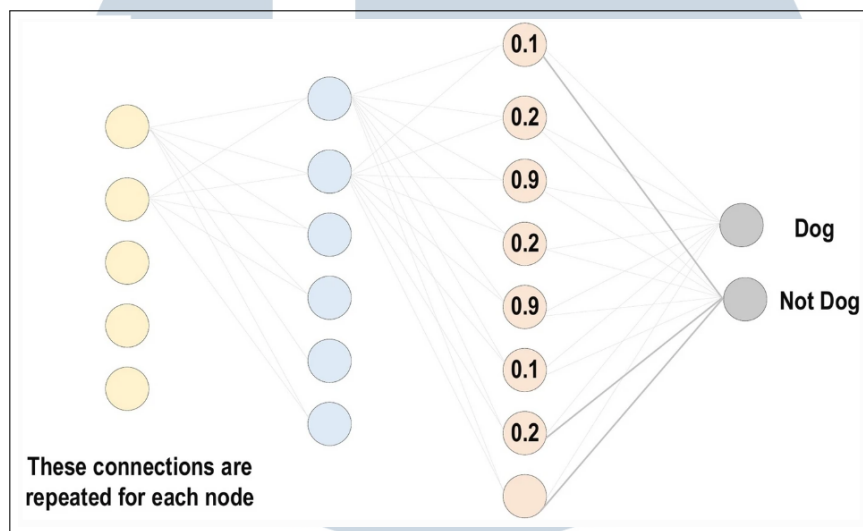


Gambar 2.2. Pooling Layer

Sumber: [11]

Pooling layer berfungsi untuk melakukan *sub-sampling* terhadap *feature map* dari proses *convolution*. Tujuannya mengubah *feature map* dari ukuran besar

menjadi lebih kecil, namun fungsi ini tetap mempertahankan informasi utama dari fitur data tersebut. Terdapat beberapa metode *pooling* yang biasa digunakan yaitu *Max Pooling*, *Min Pooling*, *Average Pooling*, *Global Average Pooling*, *Global Max Pooling*, dan *Tree Pooling*. Dari berbagai macam *pooling* *Max Pooling*, *Min Pooling*, dan *Global Average Pooling* paling sering digunakan[11]. Ketiga *Pooling* tersebut dapat dilihat pada Gambar 2.2.



Gambar 2.3. Fully Connected Layer

Sumber: [11]

Fully Connected layer merupakan layer paling akhir pada arsitektur *CNN*. lapisan ini berfungsi sebagai *Klasifikator* utama dalam *CNN*. Pada lapisan ini, setiap *neuron* terhubung secara penuh (*fully connected*) dengan seluruh *neuron* di lapisan sebelumnya. *Input* yang diterima oleh lapisan ini berasal dari *convolution layer* atau *pooling layer*. Data dalam bentuk *feature map* diubah menjadi *vektor satu dimensi* melalui proses *flattening*, vektor inilah yang menjadi *input* ke *fully connected layer*. *Output* dari layer ini berupa hasil dari *klasifikasi*, biasanya dapat berupa *probabilitas* pada setiap kelas *klasifikasi*[11]. Proses ini dapat dilihat pada Gambar 2.3.

2.3 EfficientNet

EfficientNet merupakan bagian dari arsitektur *neural network* yang dirancang untuk menyeimbangkan antara kinerja dan efisiensi. Saat ini *EfficientNet* terdiri dari beberapa model *EfficientNet-B0* hingga *EfficientNet-B7* yang merupakan

dari 16 ke 320. Perubahan ini memungkinkan model menangkap fitur lebih kompleks dari level rendah hingga tinggi.

5. MBConv 6, 3×3

Merupakan blok terakhir sebelum masuk ke proses klasifikasi.

6. Klasifikasi

Setelah melewati blok terakhir kemudian fitur diekstrak menggunakan Global Average Pooling (GAP) untuk merangkum informasi. Hasil akhir akan masuk ke Fully Connected (FC) dan softmax untuk melakukan klasifikasi gambar.

2.4 Fast Fourier Transform

Fast Fourier Transform (FFT) adalah teknik yang digunakan untuk mengubah sinyal analog menjadi sinyal digital dalam domain frekuensi. Metode ini menganalisis suatu sinyal dengan memecahnya menjadi berbagai komponen frekuensi yang direpresentasikan dalam fungsi eksponensial kompleks[15]. Dalam pemrosesan gambar, *Fast Fourier Transform (FFT)* digunakan untuk mengkonversi gambar dari domain spasial ke domain frekuensi dengan memisahnya menjadi komponen real dan imajiner. Transformasi ini memungkinkan analisis serta manipulasi gambar berdasarkan karakteristik frekuensinya[16].

$$F(x,y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{-i \times 2 \times \pi (x \frac{m}{M} + y \frac{n}{N})} \quad (2.1)$$

$$f(x,y) = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(m,n) e^{i \times 2 \times \pi (x \frac{m}{M} + y \frac{n}{N})} \quad (2.2)$$

- $f(m,n)$ adalah nilai piksel pada posisi (m,n) dalam gambar asli (domain spasial).
- $F(x,y)$ adalah representasi gambar dalam domain frekuensi, di mana koordinat (x,y) menunjukkan posisi dalam spektrum frekuensi.
- $M.N$ adalah ukuran gambar dalam piksel.
- i adalah bilangan imajiner $\sqrt{-1}$ yang digunakan dalam ekspresi eksponensial kompleks pada FFT.

2.5 Confusion Matrix

Confusion matrix adalah metode evaluasi model dalam bentuk visual. Kolom dalam *confusion matrix* menggambarkan *instance* pada kelas yg di prediksi, sedangkan baris menggambarkan *instance* yang sebenarnya[17]. *Confusion matrix* dapat dilihat pada Gambar 2.5.

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

Gambar 2.5. Confusion matrix

Sumber: [18]

Confusion matrix digunakan untuk menilai hasil prediksi dari model terhadap data yang diberikan. Penilaian tersebut meliputi *accuracy*, *precision*, *recall* dan *F1-score*. Adapun persamaan yang digunakan dalam untuk menghitung *accuracy*, *precision*, *recall* dan *F1-score* adalah sebagai berikut

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.5)$$

$$F1 - \text{Score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (2.6)$$