

BAB 2

LANDASAN TEORI

Dalam klasifikasi gambar vector buatan manusia atau AI, terdapat beberapa teori yang dipahami seperti algoritma CNN, augmentasi dan evaluasi.

2.1 Convolutional Neuron Network

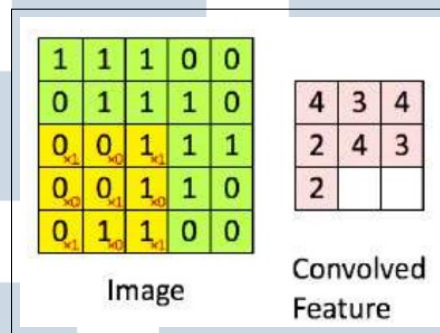
Convolutional Neural Network (CNN) merupakan jenis jaringan saraf tiruan yang sering digunakan dalam analisis gambar. Metode *deep learning* yang menggunakan *Convolutional Neural Network* (CNN) berhasil melakukan tugas klasifikasi objek pada gambar dengan lebih baik daripada metode lainnya, seperti *Support Vector Machine* (SVM), pada kompetisi pengenalan gambar skala besar *ImageNet Large Scale Visual Recognition Challenge* 2012 [11].

Tiga jenis lapisan utama pada arsitektur CNN adalah *Convolutional layer*, *pooling layer* dan *fully connected layer*. Setiap lapisan dalam CNN mengubah volume masukan tiga dimensi menjadi volume keluaran tiga dimensi yang berisi aktivasi sel saraf. Tidak seperti *Multi-Layer Perceptron* (MLP) yang arsitekturnya dua dimensi, CNN menggunakan arsitektur tiga dimensi yang mencakup lebar, tinggi, dan kedalaman. Lapisan konvolusi melakukan sebagian besar komputasi, sedangkan lapisan *pooling*, khususnya *max-pooling*, berfungsi untuk mengurangi ukuran data setelah konvolusi dengan mengurangi jumlah sampel. *Max-pooling* adalah metode *pooling* yang umum digunakan, di mana nilai maksimum dalam area tertentu dipilih.

2.1.1 Convolutional Layer

CNN utama terdiri dari lapisan konvolusi, yang melakukan fungsi konvolusi pada output dari lapisan sebelumnya [12]. Mengaplikasikan fungsi pada output fungsi lain berulang kali disebut konvolusi. Dalam pengolahan gambar, konvolusi berarti menerapkan kernel, atau kotak kuning, pada gambar di setiap sudut yang mungkin, seperti yang ditunjukkan pada Gambar 2.1. Kotak hijau secara keseluruhan adalah gambar yang akan dikonvolusi. Kernel bergerak dari sudut kiri atas ke sudut kanan bawah. Pergeseran ini menunjukkan hasilnya pada gambar di sebelah kanan.

Konvolusi digunakan untuk mengekstraksi fitur dari gambar input. Ini menghasilkan transformasi linear dari data input yang sesuai dengan informasi spasial yang ada pada data. Kernel konvolusi yang digunakan ditentukan oleh bobot lapisan ini. Input CNN dapat dilatih untuk kernel ini.



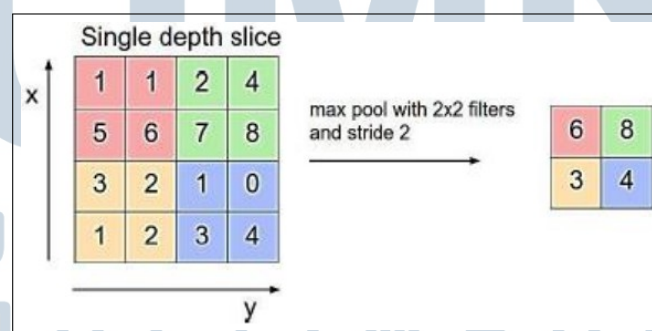
Gambar 2.1. Convolutional Layer

Sumber: [12]

2.1.2 Pooling Layer

Lapisan ini melakukan proses untuk mengurangi ukuran data gambar, yang mengurangi kesulitan komputasi dan mengontrol *overfitting*. *Max pooling* adalah salah satu dari banyak jenis lapisan campuran. Jenis ini membagi

output lapisan *convolution* menjadi bagian kecil, atau *grid*. Untuk membuat matriks gambar yang sudah direduksi, nilai maksimum diambil dari setiap *grid* yang telah dibuat, seperti yang ditunjukkan pada Gambar 2.2.



Gambar 2.2. Max Pooling

Sumber: [12]

Grid merah, hijau, kuning, dan biru menunjukkan bagian mana yang memiliki nilai tertinggi. Proses ini menghasilkan kumpulan *grid* di sebelah kanan.

Proses *max pooling* membantu memastikan bahwa fitur yang terdeteksi tetap sama meskipun objek dalam gambar berubah.

2.1.3 Fully-Connected Layer

Fully-Connected Layer adalah lapisan di mana setiap neuron aktivasi dari lapisan sebelumnya terhubung ke setiap neuron di lapisan berikutnya, mirip dengan jaringan saraf tiruan pada umumnya. *Multi-Layer Perceptron* (MLP) biasanya menggunakan lapisan ini untuk mengubah dimensi data agar dapat diklasifikasikan secara linear

Neuron dalam lapisan *Fully-Connected* terhubung ke semua input, sedangkan neuron dalam lapisan konvolusi hanya terhubung ke area tertentu. Ini adalah perbedaan utama antara lapisan konvolusi dan lapisan Fully-Connected. Tetapi fungsinya tidak jauh berbeda karena kedua lapisan tersebut melakukan operasi dot product.

2.1.4 Augmentasi Data

Augmentasi data merupakan metode yang memungkinkan praktisi untuk meningkatkan variasi data yang tersedia untuk pelatihan model secara signifikan, tanpa perlu mengumpulkan data baru secara langsung. Teknik-teknik augmentasi seperti *cropping*, *padding*, dan *horizontal flipping* biasanya digunakan dalam pelatihan jaringan *neural* yang besar. Namun, kebanyakan metode yang diterapkan dalam pelatihan jaringan *neural* masih menggunakan tipe augmentasi yang sederhana, sementara arsitektur jaringan *neural* sendiri telah diteliti secara mendalam [13].

2.1.5 Evaluasi

Proses evaluasi dilakukan untuk mengevaluasi tingkat keberhasilan model yang telah dikembangkan. Ini dilakukan dengan membandingkan kinerja model pada data pelatihan dan data pengujian dengan memanfaatkan metrik seperti akurasi, presisi, recall, dan nilai F1, yang dihitung dengan menggunakan *confusion matrix*.

Dalam *machine learning*, *confusion matrix* banyak digunakan untuk klasifikasi yang diawasi atau untuk menganalisis kinerja model klasifikasi. Struktur *confusion matrix* berbentuk persegi dengan baris dan kolom, yang menunjukkan

kelas data aktual dan diprediksi. *confusion matrix* berbentuk matriks 2×2 untuk klasifikasi biner. Empat komponen utama terdiri dari matriks ini yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Untuk masalah klasifikasi multikelas, *confusion matrix* berukuran $k \times k$, dengan k sebagai jumlah kelas [14].

Tabel 2.1. Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)
Aktual Negatif	False Positive (FP)	True Negative (TN)

Presisi (P) adalah rasio antara prediksi positif yang benar dengan seluruh prediksi positif [15], yang dirumuskan pada persamaan 2.1.

$$P = \frac{TP}{TP + FP} \times 100\% \quad (2.1)$$

Recall (R) adalah rasio antara prediksi positif yang benar dengan seluruh sampel positif aktual [15], yang dirumuskan pada persamaan 2.2.

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2.2)$$

F1-score merupakan rata-rata harmonis dari presisi dan recall yang digunakan untuk mengukur performa model [15], yang dirumuskan pada persamaan 2.3

$$F1\text{-score} = \frac{2 \times P \times R}{P + R} \quad (2.3)$$

Akurasi adalah persentase data yang berhasil diklasifikasikan dengan benar dari seluruh data yang diuji, dan merupakan metrik performa yang paling intuitif [15] dirumuskan pada persamaan 2.4.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.4)$$

2.1.6 Grad-CAM

Grad-CAM adalah sebuah teknik yang berguna untuk menunjukkan area mana yang diperhatikan oleh model *CNN* saat melakukan prediksi. Teknik ini dapat membantu menjelaskan hasil keluaran dari sebuah model *machine learning*. Grad-CAM memanfaatkan informasi *gradients* dari *convolutional layer* terakhir untuk membuat peta lokasi yang menandai bagian-bagian penting pada citra input yang berkontribusi terhadap prediksi. Selain menjelaskan aktivasi pada *convolutional layer* terakhir, teknik ini juga dapat diterapkan pada seluruh *convolutional layers*, sehingga memungkinkan *explainability* dilakukan secara menyeluruh [16].

2.1.7 Penelitian Terdahulu

Tabel 2.2. Penelitian Terdahulu

Judul Artikel	Jurnal	Penulis	Hasil dan Kesimpulan
Detection of AI-Generated Anime Images Using Deep Learning[8]	ICIC Express Letters (2024)	Surya Widi Kusuma, Friska Natalia, Chang Seong Ko, Sud Sudirman	Penelitian ini membandingkan arsitektur <i>MobileNetV2</i> dan <i>MobileNetV3</i> dalam mendeteksi gambar anime buatan AI menggunakan <i>transfer learning</i> . Hasil menunjukkan <i>MobileNetV3</i> lebih unggul dengan akurasi 97%, presisi 100%, dan recall 95%, serta kemampuan generalisasi yang baik.
Klasifikasi Gambar Pemandangan dengan Kecerdasan Buatan Berbasis CNN [9]	Jurnal JTIK (2024)	Mozes Hasian Veltin Sinaga, Muhammad Albirra, Muhammad Fajar Sidiq	Penelitian ini menggunakan CNN untuk mengklasifikasikan gambar pemandangan seperti gunung, hutan, jalan, laut, dan gletser. Model berhasil mengenali fitur visual dari tiap kategori dan menghasilkan akurasi yang memuaskan dalam klasifikasi gambar pemandangan baru.
Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode Convolutional Neural Network [10]	TEKNIKA (2024)	Jesselyn Mu, Muhammad Adrezo, Ahmed Nizhan Haikal	Penelitian ini menggunakan CNN dengan 6 convolutional layer dan batch normalization untuk membedakan wajah asli dan <i>deepfake</i> . Model dilatih dengan dataset Kaggle dan menunjukkan akurasi, presisi, recall, dan F1-score sebesar 91%, membuktikan efektivitas CNN dalam deteksi gambar deepfake.

Berdasarkan penelitian-penelitian terdahulu yang dirangkum dalam Tabel 2.2, dapat disimpulkan bahwa metode *Convolutional Neural Network* (CNN) telah banyak digunakan dalam proses klasifikasi gambar. Artikel “Detection

of AI-Generated Anime Images Using Deep Learning” [8] menunjukkan bahwa pendekatan CNN berbasis arsitektur *MobileNetV3* mampu mendeteksi gambar anime buatan AI dengan akurasi tinggi, yaitu mencapai 97%. Artikel “Identifikasi Wajah Asli dan Buatan Deepfake Menggunakan Metode CNN” [10] juga memperlihatkan bahwa CNN dapat mengidentifikasi wajah *deepfake* secara efektif dengan F1-score sebesar 91

Sementara itu, artikel “Klasifikasi Gambar Pemandangan dengan Kecerdasan Buatan Berbasis CNN” [2] memperlihatkan bahwa CNN juga efektif digunakan untuk klasifikasi gambar umum seperti pemandangan alam, dengan performa akurasi yang memuaskan. Dari ketiga penelitian tersebut, dapat dilihat bahwa CNN memiliki kemampuan generalisasi yang baik untuk mengenali pola visual kompleks pada gambar buatan manusia maupun AI. Model CNN yang dirancang akan digunakan sebagai *binary classifier* untuk membedakan gambar kartun hewan yang dibuat oleh manusia dan yang dihasilkan oleh model AI generatif seperti Gemini AI.

