PENGENALAN AKTIVITAS MANUSIA PADA SUPERMARKET MENGGUNAKAN OPENPOSE DAN CNN

Alvian Wijaya [1)] Lina Lina[2)]

[1) 2)] Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta
email : alvian.535200026@stu.untar.ac.id[1)] ,lina@fti.untar.ac.id[2)]

ABSTRACT

Human activity recognition is a dynamic area within artificial intelligence. It involves identifying human actions during everyday tasks such as standing, sitting, and walking. One application of this technology is in supermarkets, where it can analyze consumer behavior or function as a surveillance tool to prevent theft. This particular study utilizes OpenPose and Convolutional Neural Networks (CNN) with a custom-collected dataset. The program detects human skeleton shapes from camera footage and classifies these shapes using CNN with the ResNet50 model, subsequently displaying the identified activities. The classified activities include standing, walking, picking up items, looking at items, and pushing a trolley. The testing results indicate a training accuracy of 99.76% and a validation accuracy of 96.52%, along with an accuracy score of 96.52%, a precision of 96.59%, a recall of 96.525, and an F1-score of 96.53%.

Key words

Convolutional Neural Network, OpenPose, Human Activity Recognition, ResNet50, Supermarket

1. Pendahuluan

Teknologi kecerdasan buatan saat ini telah berkembang pesat di berbagai aspek kehidupan manusia. Kecerdasan buatan dikembangkan untuk menciptakan sistem yang memiliki kemampuan mirip dengan kecerdasan manusia, seperti memahami perintah, mengenali gambar, dan memantau aktivitas. Salah satu cabang ilmu dari kecerdasan buatan adalah computer vision. Computer vision sering digunakan dalam pengolahan citra karena memiliki kemampuan untuk mengenali objek secara otomatis [1].

Salah satu contoh penerapan kecerdasan buatan dalam kehidupan manusia adalah pengenalan aktivitas manusia. Pengenalan aktivitas manusia ini melibatkan klasifikasi aktivitas berdasarkan rekaman sensor yang menangkap gerakan manusia [2]. Pengenalan aktivitas manusia dapat diterapkan dalam berbagai kegiatan, seperti pengawasan, analisis aktivitas, dan penghitungan keramaian suatu lokasi. Pengembangan teknologi ini bisa dilakukan di berbagai tempat, seperti ruang kelas, tempat umum, dan supermarket.

Supermarket pertama kali hadir di Indonesia sekitar tahun 1970, dan pada akhir 1990-an, supermarket dengan merek internasional mulai masuk ke Indonesia [3]. Supermarket, atau toko swalayan, adalah toko yang menjual berbagai kebutuhan sehari-hari seperti pangan, kebutuhan rumah tangga, pakaian, dan lainnya, tergantung kebijakan masing-masing supermarket. Untuk mendukung keamanan, supermarket biasanya memasang kamera pengawas di beberapa titik untuk memantau aktivitas selama berbelanja. Namun, penggunaan metode konvensional seperti pengawasan oleh tenaga manusia atau kamera pengawas saja masih belum cukup efektif untuk mendukung kinerja supermarket.

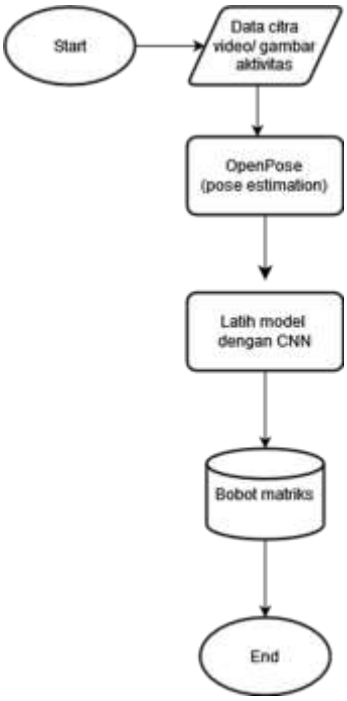
Situasi ini mendorong penerapan dan pengembangan teknologi di lingkungan supermarket, salah satunya adalah Amazon Go. Amazon Go adalah contoh implementasi pengenalan aktivitas manusia, computer vision, dan algoritma deep learning. Teknologi "Just Walk Out" di Amazon Go mencatat barang yang dibeli, dikembalikan ke rak, hingga proses checkout secara otomatis [4]. Pengenalan aktivitas manusia di supermarket menggunakan OpenPose dan Convolutional Neural Networks adalah salah satu bentuk pengembangan kecerdasan buatan dalam kehidupan sehari-hari. OpenPose adalah sistem open source yang digunakan untuk mendeteksi kerangka tubuh manusia dari sendi atau bagian tubuh dalam gambar atau video berbasis Convolutional Neural Networks secara real-time [5]. Dengan membuat titik-titik kerangka dari hasil deteksi OpenPose dari bagian tubuh bawah hingga atas, data tersebut akan digunakan sebagai data latih dengan metode Convolutional Neural Networks (CNN) menggunakan arsitektur ResNet50. Metode CNN dengan arsitektur ResNet50 ini akan membangun model yang mampu mengenali aktivitas yang sedang dilakukan berdasarkan bentuk kerangka yang terdeteksi. OpenPose dipilih karena kemampuannya melakukan pendeteksian multiperson dalam satu frame dan keunggulannya dibandingkan metode pose lainnya dalam menampilkan titik keypoint dari gerakan video dengan akurasi tinggi [6].

Diharapkan nantinya sistem yang dirancang dengan mengimplementasikan Human Activity Recognition dapat mengidentifikasi aktivitas manusia seperti mengambil barang, mengamati produk, mendorong troli, berjalan, dan berhenti untuk melihat. Aktivitas yang diawasi secara real-time ini dapat digunakan sebagai kamera pengawas yang mampu mengenali aktivitas yang sedang dilakukan, serta sebagai teknologi layanan mandiri seperti yang digunakan di Amazon Go [7].

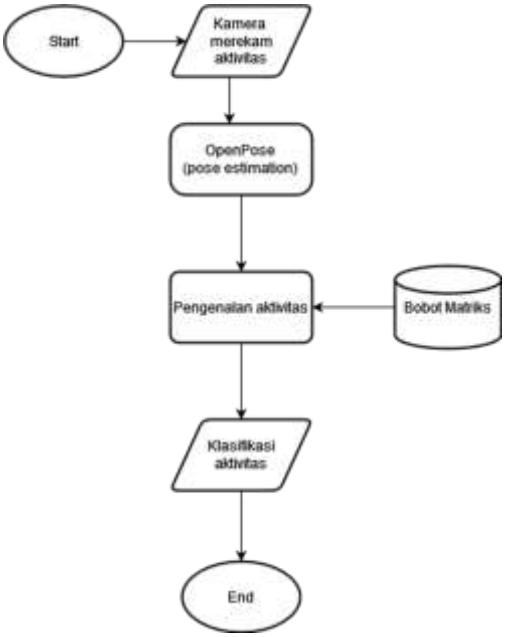
2. Metode Penelitian

Rancangan sistem yang dibuat adalah sebuah program desktop yang dapat melakukan deteksi aktivitas manusia di supermarket dengan metode OpenPose dan CNN ResNet-50.

Aplikasi dibangun secara desktop dan pelatihan dilakukan dengan menggunakan google collabs untuk mendapatkan hasil latih model CNN yang akan dipakai nantinya dalam tahapan rancangan.



Gambar 1. *Flowchart* Pelatihan



Gambar 2. *Flowchart* Tahapan Rancangan

2.1 Dataset

Data yang dikumpulkan sesuai dengan jumlah kelas aktivitas yang ada terdiri atas berdiri, berjalan, mendorong troly, mengambil barang, dan melihat barang. Sumber data didapatkan dengan cara melakukan pengambilan secara langsung menggunakan kamera langsung di supermarket maupun tempat umum lainnya. Seluruh data yang dikumpulkan dijadikan satu menjadi dataset untuk penelitian ini.

Tabel 1. Jumlah data yang dipakai

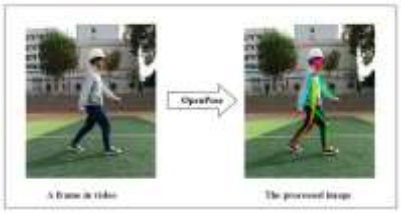
Jenis Data		Jumlah Data	Data Keseluruhan
Pelatihan Model	Berjalan	750	3.679
	Berdiri	761	
	Mendorong Troly	750	
	Mengambil Barang	700	
	Melihat barang	718	
Pengujian Model	Berjalan	206	920
	Berdiri	194	
	Mendorong Troly	150	
	Mengambil Barang	176	
	Melihat barang	194	
Total			4.599

Setelah dataset terkumpul selanjutnya akan dilakukan ekstrasi dari titik OpenPose berupa bagian dari badan dan tangan dari titik *skeleton* yang ada dengan menjalankan fungsi untuk ekstrasi titik tersebut dan akan

didapatkan 3 hasil berupa .npy untuk badan dan tangan dan jpg untuk titik *skeleton*.

2.2 OpenPose

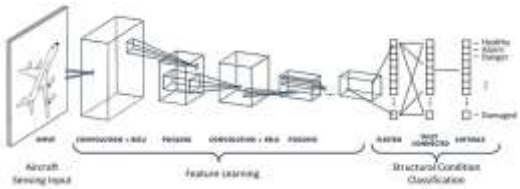
OpenPose adalah sebuah *program open source* yang dikembangkan oleh *Carnegie Mellon University (CMU).* Program ini berbasis *Convolutional Neural Networks (CNN)* dan dibangun pada *framework Caffe (Convolutional Architecture for Fast Feature Embedding)* [5]. Jumlah keluaran titik-titik kerangka dari *OpenPose* berjumlah total 25 titik dengan posisi pada sumbu x, y, dan z. *OpenPose* menggunakan pendekatan bottom-up sebagai metode penyelesaiannya [8]. OpenPose dipilih karena dapat melakukan *multiperson detection* dalam satu *frame* dan *OpenPose* lebih baik dibandingkan metode pose lainya dalam menampilkan titik keypoint dari gerakan video dengan akurasi yang tinggi [6]. Gambar 2.8 Contoh citra setelah dideteksi *OpenPose*.



Gambar 3. Visualisasi gambar dideteksi *OpenPose*

2.3 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) adalah sebuah struktur yang meniru jaringan saraf tiruan, sering digunakan untuk memproses data citra atau gambar [9]. Arsitektur CNN memiliki 3 jenis layer utama dalam prosesnya, yaitu *Convolutional Layer*, *Pooling Layer*, dan *Fully Connected Layer*. Berikut adalah gambaran proses CNN diilustrasikan dalam Gambar 4.



Gambar 4. Arsitektur CNN

Dalam Convolutional Layer, yang merupakan lapisan pertama dalam CNN, fitur-fitur diekstraksi dari gambar input.

Convolutional Layer sebagai lapisan pertama dalam CNN dalam melakukan ekstraksi fitur. Gambar input berupa matrik akan dihitung dengan persamaan seperti berikut:

$$x(i,j) = \sum m \sum n\, w^{l}_{m,n} * o^{l-1}_{i+m,j+n} + b \quad (1)$$

Keterangan:
x(i,j) = hasil perhitungan konvolusi pada titik (x, y)
l = layer
o_(i,j)= input citra
w_(m,n) = filter yang dipakai
b = bias
i = baris pixel citra
j = kolom pixel citra

Selanjutnya feature map untuk menghitung matrik hasil convolution dengan rumus:

$$Output = \frac{w-n}{s} + 1 \quad (2)$$

Keterangan:
w = Panjang atau tinggi input
n = Panjang atau tinggi filter
p = Padding
s = Stride

Kemudian dengan aktivasi ReLU untuk memilah pixel yang kurang dari 0 (nol) lalu mengubah nilainya menjadi 0 (nol) dengan rumus:

$$f(x) = \max(0,x) \quad (3)$$

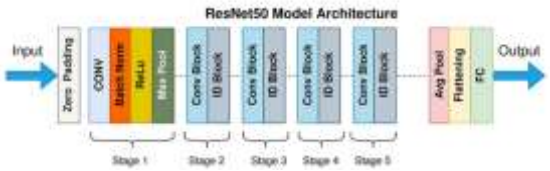
Pooling Layer berfungsi untuk mengurangi dimensi gambar dengan memilih nilai yang penting dari setiap wilayah. Teknik-teknik umum dalam *Pooling Layer* meliputi *Max Pooling*, *Average Pooling*, *Min Pooling*, dan *Global Average Pooling*.

Flattening adalah proses mengubah matriks dua dimensi menjadi sebuah vektor.

Fully Connected Layer terdiri dari input layer, hidden layer, dan output layer. Pada input layer, dilakukan penggabungan dari seluruh matriks *feature map* yang dihasilkan pada proses pooling. Kemudian, piksel-piksel tersebut diubah menjadi vektor dengan panjang yang sama dengan nilai *pooling* [10]. Kemudian *hidden layer* digunakan untuk menghitung nilai dari input layer dengan nilai yang sudah diinisialisasikan sebelumnya [10]. Output layer adalah nilai-nilai dari *hidden layer* yang kemudian dimasukkan ke dalam fungsi aktivasi, seperti fungsi *softmax*, untuk mendapatkan nilai-nilai aktivasi dari output [10].

2.3 ResNet-50

ResNet-50 merupakan sebuah arsitektur bagian dari ResNet (Residual Network) yang memiliki jumlah 50 lapisan dalam jaringannya [11].

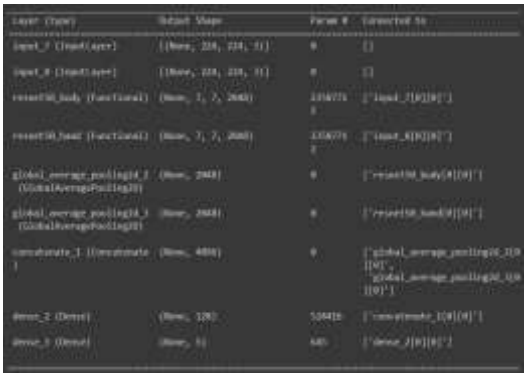


Gambar 5. Arsitektur ResNet-50

Fitur utama dari ResNet-50 adalah jumlah lapisannya yang besar, mencapai 50 lapisan. Keunggulan lainnya adalah penggunaan blok residu, yang memungkinkan aliran informasi dari satu lapisan ke lapisan berikutnya. Ini membantu mengatasi masalah dalam pelatihan model, seperti masalah penurunan kinerja yang sering terjadi pada jaringan yang lebih dalam [11].

3. Hasil Percobaan

Setelah rancangan dan pembuatan diimplementasikan selanjutnya adalah dengan membuat rancangan model program dengan *OpenPose* dan *ResNet-50*. Dari dataset yang ada akan dibuatkan *masking* titik *skeleton* lalu baru dimasukkan kedalam Model ResNet-50.



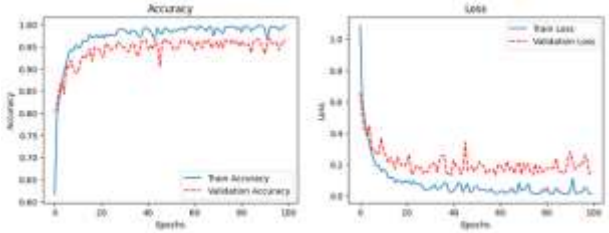
Gambar 6. Lapisan layer ResNet-50

Selanjutnya model akan divariasikan untuk menentukan model manakah yang terbaik diantara ketiga model yang dibuat. Model yang dilatih terdiri dari Batch size, Epoch, Akurasih latih, Loss latih, Akurasi validasi, dan Loss validasi.

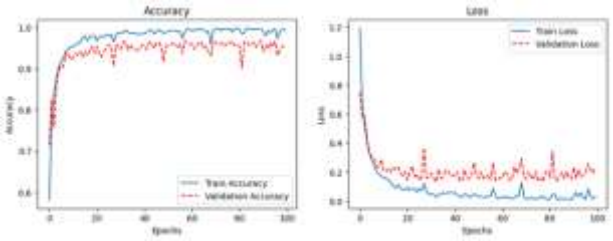
Tabel 2. Hasil Pelatihan Model

Batch size	Epoch	Akurasi latih	Loss latih	Akurasi validasi	Loss validasi
32	100	99.76	0.0075	96.52	0.1489
64	100	99.35	0.0295	95.43	0.2286
128	100	99.84	0.092	95.98	0.1472

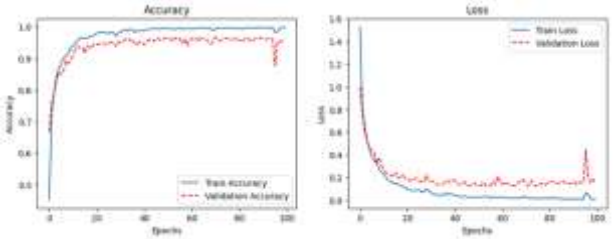
Selanjutnya setelah didapatkan hasil diatas maka dibuatkan hasil *Grafik Accuracy* dan *Grafik Loss* Model dari masing-masing model yang ada sebagai berikut.



Gambar 7. Hasil Grafik *Accuracy* dan Grafik *Loss* Model 1



Gambar 8. Hasil Grafik *Accuracy* dan Grafik *Loss* Model 2



Gambar 9. Hasil Grafik Accuracy dan Grafik Loss Model 3

Selanjutnya model dilakukan pengujian dengan data yang belum pernah di latih sebelumnya berjumlah 100 video yang dibagi ke dalam 5 kelas aktivitas kemudian didapatkan pencatatan hasilnya sebagai berikut

Tabel 3. Hasil Pengujian Model

No.	Aktivitas	Jumlah video / yang benar	Akurasi
1	Berdiri	20/14	70%
2	Berjalan	20/16	80%
3	Mengambil barang	20/15	75%
4	Mendorong troly	20/18	90%
5	Melihat barang	20/11	55%

ditunjukan bahwa akurasi tiap aktivitas mulai dari berdiri sebesar 70%, berjalan sebesar 80%, mendorong troly sebesar 75%, mengambil barang sebesar 90%, dan melihat barang sebesar 55% dari total masing-masing data uji sebanyak 20 gambar tiap kelasnya.

4. Kesimpulan

Setelah dilakukan penelitian tentang pendeteksian aktivitas manusia di supermarket, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Keseluruhan modul berjalan dengan baik, dengan fitur-fitur dan tombol yang berfungsi sesuai dengan rancangan yang telah direncanakan.
2. Program berhasil mengenali dan mengklasifikasikan aktivitas manusia, termasuk berjalan, berdiri, mendorong troli, mengambil barang, dan melihat barang.
3. Hasil pengujian menunjukkan tingkat akurasi yang bervariasi untuk setiap aktivitas, dengan akurasi tertinggi untuk mengambil barang dan

akurasi terendah untuk melihat barang, dengan total akurasi sebesar 74%.

Setelah dilakukab penelitian, beberapa saran yang dapat diterapkan di penelitian selanjutnya adalah sebagai berikut:

1. Mencoba menggunakan metode-metode alternatif untuk meningkatkan komputasi dan akurasi.
2. Menambah jumlah kelas aktivitas dan variasi dataset untuk meningkatkan keberagaman dan ketepatan hasil klasifikasi.
3. Menggunakan dataset yang lebih spesifik yang merefleksikan kondisi sebenarnya di supermarket, serta mempertimbangkan pengenalan aktivitas dari lebih dari satu subjek manusia.
4. Melakukan eksplorasi terhadap arsitektur CNN lainnya untuk membandingkan hasil dan kinerjanya.

REFERENSI

[1] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat, "CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope," *Electronics*, vol. 10, no. 20, p. 2470, 2021.

[2] R. R. Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, vol. 19, no. 2, pp. 302-311, 2020.

[3] D. Suryadarma, A. Poesoro, S. Budiyati, Akhamadi, and M. Rosfadhila, "Dampak Supermarket terhadap Pasar dan Pedagang Ritel Tradisional di Daerah Perkotaan di Indonesia," SMERU, Jakarta, 2007.

[4] A. Polacco and K. Backes, "The Amazon Go Concept: Implications, Applications, and Sustainability," *Journal of Business and Management*, vol. 24, no. 1, pp. 79-92, 2018.

[5] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172-186, 2019.

[6] S. Mroz, N. Baddour, C. J. C. McGuirk, and P. Juneau, "Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose," dalam *4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, Paris, 2021.

[7] S. Junsawang, W. Chaiyasoonthorn, and S. Chaveesuk, "Willingness to Use Self-Service Technologies Similar to Amazon Go at Supermarkets in Thailand," dalam *MSIE 2020: 2020 2nd International Conference on Management Science and Industrial Engineering*, Osaka, 2020.

[8] K. D. Runtu and Lina, "Pengenalan Aktivitas Manusia di Supermarket dengan Metode Long Short Term Memory," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 10, no. 2, 2022.

[9] M. Arsal, B. A. Wardijono, and D. Anggraini, " Face Recognition Untuk Akses Pegawai Bank Menggunakan Deep Learning Dengan Metode CNN," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 1, pp. 55-63, 2020.

[10] F. M. Qotrunnada and P. H. Utomo, "Metode Convolutional Neural Network untuk Klasifikasi Wajah Bermasker," *Prosiding Seminar Nasional Matematika*, vol. 5, pp. 779-807, 2022.

[11] S. Mukherjee, "The Annotated ResNet-50," Toward Data Science, 18 August 2022. [Online]. Available: https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758. [Diakses 11 September 2023].

Alvian Wijaya, Mahasiswa S1, program studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara.

Prof. Lina, S.T, M. Kom., Ph.D., Memperoleh gelar Sarjana dari Universitas Tarumanagara, Indonesia tahun 2001 dan gelar Magister dari Universitas Indonesia, Indonesia tahun 2004, kemudian ditahun 2009 memperoleh gelar Ph.D dari Nagoya University, Jepang. Saat ini sebagai Dosen tetap Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara.