

## Navigasi Menu Berdasarkan Arah Pandangan Mata pada Kursi Roda Pintar menggunakan Fusion-CNN berbasis Jetson TX2

Blessius Sheldo Putra Laksono<sup>1</sup>, Fitri Utamingrum<sup>2</sup>

Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>blessiussheldo@student.ub.ac.id, <sup>2</sup>f3\_ningrum@ub.ac.id

### Abstrak

Teknologi saat ini telah mengalami kemajuan yang pesat, tetapi interaksi antara komputer dan manusia belum mengalami kemajuan yang sama. Ini merupakan masalah bagi pengguna teknologi yang mengalami disabilitas gerak. Kursi roda pintar telah banyak digunakan untuk membantu orang dengan disabilitas, tetapi masalah interaksi dengan komputer masih mengganggu kenyamanan pengguna. Penelitian sebelumnya telah menawarkan beberapa metode interaksi baru, seperti menggunakan suara, tetapi metode ini dianggap kurang efektif karena membutuhkan lingkungan yang minim noise. Interaksi lain yang ditawarkan adalah dengan menggunakan estimasi arah pandangan mata dengan menggunakan algoritma konvensional atau machine learning. Namun, algoritma konvensional dianggap kurang efektif karena adaptabilitasnya yang rendah dengan pengguna yang berbeda. Algoritma CNN dipilih dalam penelitian ini karena kemampuan untuk mengambil fitur dari citra, sehingga algoritma dapat beradaptasi dengan data baru. Penelitian ini menunjukkan bahwa akurasi model sebesar 96% dengan loss sebesar 0.02 pada fase training. Sistem dapat menjalankan algoritma ini dalam waktu 0.16 detik menggunakan akselerasi CUDA. Sistem hanya menggunakan daya listrik sebesar 12 Watt, yang memungkinkan sistem dijalankan dengan baterai. Hasil pengujian menunjukkan bahwa sistem dapat berjalan dengan baik dalam melakukan estimasi arah pengguna.

**Kata kunci:** *Fusion CNN, Computer vision, Kursi roda pintar, Kecerdasan buatan, Gaze estimation*

### Abstract

*As time progresses, technology also experiences rapid advancement. However, this advancement is not accompanied by an improvement in the interaction between computers and humans. This is certainly a significant impact on users of technology, particularly those with mobility disabilities. Smart wheelchair technology has already been widely used to assist people with disabilities, but the problem of interaction with computers still reduces user comfort. Some new types of interaction have been offered in previous research, including using voice, but this method is considered less effective due to the need for a minimally noisy environment. Another offered method of interaction is using eye gaze estimation using conventional algorithms or machine learning. For this, conventional algorithms are considered less effective due to their low adaptability with different users. CNN-based algorithms are chosen in this research because of their ability to extract features from images, allowing the algorithm to adapt to new data. The accuracy of the model in this research was 96% with a loss of 0.02 during the training phase. The system can run the algorithm in 0.16 seconds using CUDA acceleration. The system only uses 12 watts of electricity, making it possible to run the system using a battery. From the testing that was carried out and the results obtained, it can be concluded that the system runs well to estimate the direction of the user.*

**Keywords:** *Fusion CNN, Computer vision, Smart wheelchair, Artificial intelligence, Gaze estimation*

### 1. PENDAHULUAN

dengan kemajuan zaman membuat penggunaan teknologi menjadi sangat penting, bahkan menjadi kecakapan yang harus dikuasai.

Penggunaan teknologi saat ini tidak hanya terfokus pada pekerjaan-pekerjaan khusus di bidang tertentu saja, tapi juga menjadi kebutuhan sehari-hari bagi banyak orang. Sayangnya, kemajuan teknologi yang pesat tidak diikuti dengan kemudahan akses bagi orang dengan

disabilitas fisik. Disabilitas fisik yang dimaksud adalah gangguan fungsi gerak seperti lumpuh atau kaku terutama pada tangan. Saat ini, orang harus menggunakan tangannya untuk memberikan input ke komputer, seperti menggunakan layar sentuh atau papan ketik. Namun, orang dengan disabilitas fisik terutama pada tangan tidak dapat menggunakan cara interaksi tersebut

Salah satu alat bantu yang sering digunakan oleh orang dengan disabilitas fisik adalah kursi roda. Ada banyak jenis kursi roda pintar yang dikembangkan dengan fitur-fitur yang membantu orang dengan disabilitas fisik, seperti pengendalian dengan suara (Rockland & Reisman, 1998) (Anam & Saleh, 2020), pengendalian dengan arah pandangan (Araujo et al., 2020) (Pangestu & Utaminigrum, 2020), dan pengendalian dengan pergerakan otot pengguna (Aihara et al., 2022). Penelitian sebelumnya mengenai pengembangan kursi roda pintar menunjukkan bahwa pengguna kursi roda harus berinteraksi dengan menu pada display menggunakan layar sentuh. Hal ini tentu tidak mudah bagi orang dengan disabilitas ganda yang memiliki keterbatasan pada tangan dan mempersulit penyandang disabilitas fisik dalam memilih menu yang ada pada layar.

Beberapa cara interaksi lain telah diusulkan sebelumnya seperti menggunakan suara atau speech detection untuk mengendalikan menu pada layar, namun cara ini dinilai kurang efektif terutama dalam lingkungan yang bising. Salah satu cara interaksi yang paling efektif adalah dengan menggunakan gaze estimation, yaitu memperkirakan arah pandang pengguna terhadap antarmuka pada layar. Gaze estimation dapat menggunakan beberapa metode seperti elektroda pada kulit sekitar bagian mata (electrooculography), kamera infra merah, atau kamera RGB.

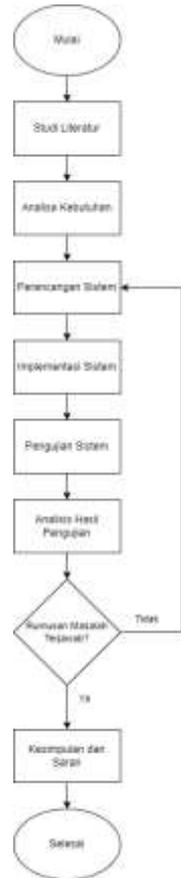
Metode EOG melakukan pengukuran sinyal potensial elektrik antara dua titik pada sekitar mata. Pergerakan otot mata menghasilkan sinyal elektrik yang mengubah pembacaan EOG. Empat titik pengukuran EOG diperlukan untuk estimasi pandangan pengguna, dua untuk mengukur pergerakan horizontal dan dua untuk pergerakan vertikal. Namun, penggunaan EOG dapat menyebabkan ketidaknyamanan dan membutuhkan perawatan tambahan, serta harus diaplikasikan oleh orang lain pada saat pertama kali digunakan. Alternatif lain adalah menggunakan kamera untuk gaze estimation pada video stream yang dihasilkan. Gaze

estimation dapat dilakukan dengan kamera IR dan sumber cahaya infra merah. Sumber cahaya ini menghasilkan pantulan (glint) pada mata pengguna yang membantu sistem menentukan arah pandangan relatif terhadap kamera (Mohan & Phirke, 2020). Hasil yang lebih optimal dapat dicapai dengan menggunakan lebih dari satu sumber cahaya dengan posisi yang berbeda (Yoon et al., 2019)..

Penggunaan kamera RGB dapat memperkecil biaya dan bahkan dapat menggunakan kamera RGB di pasaran tanpa modifikasi (Akinyelu & Blignaut, 2022). Ada 2 jenis metode yang biasa digunakan dengan kamera RGB, yaitu metode konvensional dan metode deep learning. Metode konvensional menggunakan fitur dari gambar wajah pengguna untuk menentukan arah pandangan, sementara metode appearance based dengan deep learning saat ini lebih banyak digunakan karena memberikan hasil yang lebih akurat (Fischer et al., 2018). Pemilihan arsitektur neural network yang tepat menjadi faktor penting untuk mendapatkan model yang akurat dan efektif dalam mendeteksi arah pandangan. Berdasarkan studi literatur yang telah dilakukan, peneliti mengusulkan judul penelitian Navigasi Menu Berdasarkan Arah Pandangan Mata pada Kursi Roda Pintar menggunakan Fusion-CNN Berbasis Jetson TX2.

## 2. METODOLOGI

Pada penelitian ini digunakan tipe penelitian implementatif pengembangan, yang merupakan penelitian yang menggunakan metode penerapan perangkat lunak dan keras serta tidak hanya melakukan analisis data. Penelitian ini merupakan pengembangan dari masalah dan solusi yang diperoleh dari kekurangan penelitian sebelumnya.



Gambar 1 Tahapan Metodologi Penelitian

### 2.1 Teknik Pengumpulan dan Analisis Data

Pada penelitian ini, teknik pengumpulan data yang digunakan adalah dengan menggunakan kamera webcam Logitech C920 dan subjek yang duduk pada kursi roda pintar menghadap layar. Dataset yang digunakan terdiri dari 10000 citra yang terbagi menjadi 5000 citra mata dan 5000 citra wajah, yang kemudian dibagi lagi menjadi 5 kelas yaitu "kanan atas", "kiri atas", "kanan bawah", "kiri bawah", dan "unknown". Data akan dibagi menjadi 2 bagian, dengan 80% digunakan untuk proses training dan 20% untuk proses testing. Data untuk masing masing kelas ditunjukkan pada Gambar 2.

Data ini kemudian akan dianalisis dengan tujuan untuk memastikan model yang dihasilkan dapat memenuhi kebutuhan fungsional yang ada pada sistem. Data ini juga akan digunakan untuk melakukan *testing* terhadap model yang sudah dibentuk untuk mengetahui performa model.



Gambar 2 Contoh Dataset

## 3. REKAYASA KEBUTUHAN

### 3.1 Gambaran Umum Sistem

Sistem adalah bagian (sub-sistem) dari kursi roda pintar yang digunakan untuk memilih menu yang ada pada layar kursi roda pintar menggunakan arah pandangan mata pengguna. Menu yang dipilih adalah daftar dari fitur yang dapat diaktifkan ataupun dinonaktifkan pada sistem kursi roda pintar. Untuk melakukan estimasi arah pandangan digunakan fitur dari wajah dan mata kanan pengguna untuk mendapatkan hasil estimasi yang lebih akurat. Citra mata yang ditangkap sistem akan dilakukan *resize* menjadi ukuran  $100 \times 50$  dan citra wajah di *resize* menjadi ukuran  $100 \times 100$ . Kedua citra ini akan diubah menjadi *grayscale* untuk mengurangi beban komputasi saat konvolusi. Kemudian proses estimasi arah pandangan akan menggunakan algoritme *Fusion CNN*. Hasil estimasi arah pandangan akan ditunjukkan pada layar pengguna dengan merubah warna tombol menjadi warna kuning.

### 3.2 Rekayasa Kebutuhan Fungsional

Berdasarkan gambaran umum sistem yang dimiliki maka dapat diberikan kebutuhan fungsional sistem sebagai berikut,

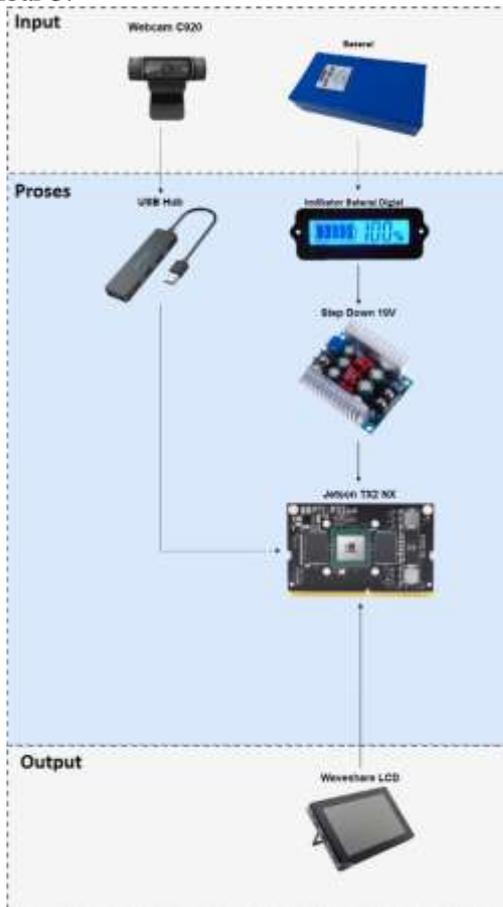
1. Sistem dapat menjalankan algoritme *facial landmark* untuk menemukan wajah pengguna.
2. Sistem dapat menunjukkan antarmuka grafik untuk pengguna kursi roda pintar dan memberi tahu pengguna hasil deteksi arah pandangan melalui antarmuka grafik tersebut.

3. Sistem dapat mendeteksi apabila pengguna dengan sengaja mengedipkan mata untuk memilih menu yang ditunjukkan pada layar.
4. Sistem dapat mengklasifikasikan citra mata dan wajah untuk menentukan arah pandangan pengguna.
5. Sistem dapat menangkap citra wajah dan mata pengguna.

#### 4. PERANCANGAN DAN IMPLEMENTASI

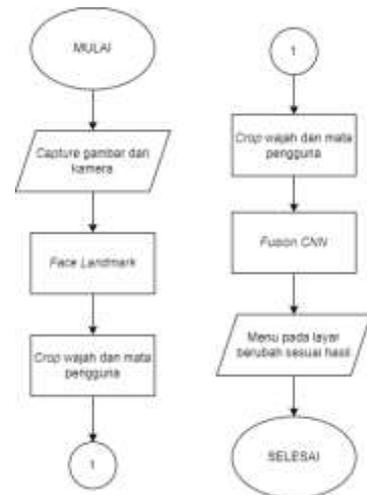
##### 4.1 Perancangan Perangkat Keras

Perangkat keras yang digunakan pada sistem ini terdiri dari beberapa komponen, ada komponen yang digunakan sebagai input dari sistem yaitu kamera Logitech C920, komponen yang digunakan sebagai output sistem yaitu layar, dan komponen yang dilakukan untuk memproses data inputan yaitu Jetson TX2. Skematik lengkap dari sistem digambarkan pada Gambar 3.



Gambar 3 Skematik Perangkat Keras Sistem

##### 4.2 Perancangan Perangkat Lunak



Gambar 4 Flowchart Program Utama

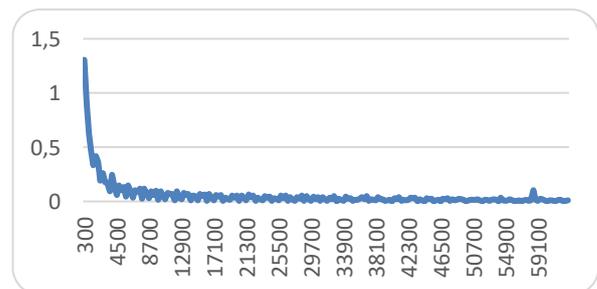
Perangkat lunak yang dirancang pada sistem ini meliputi bagian *preprocessing* dengan menggunakan *face landmark*. Kemudian hasil *preprocessing* akan diproses pada *Fusion CNN* untuk menghasilkan estimasi arah pandangan pengguna dan menampilkan menu yang dilihat pengguna pada layar.

*Preprocessing* pada sistem ini terdiri dari deteksi wajah menggunakan *face landmark* dan *cropping* pada bagian wajah dan mata. Kemudian hasil *cropping* tersebut akan diubah menjadi grayscale yang kemudian menjadi input untuk proses estimasi arah pandangan.

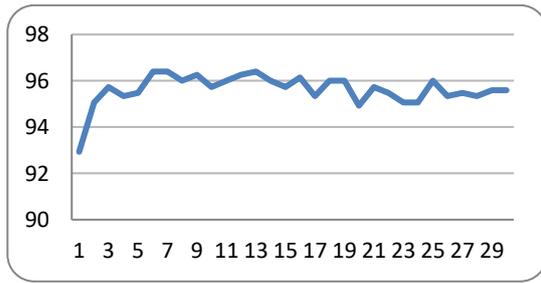
#### 5. PENGUJIAN DAN ANALISIS

##### 5.1 Pengujian Pengaruh Epoch Terhadap Performa Model

Tujuan dari pengujian ini adalah emengetahui pengaruh *epoch* yang dilakukan ketika proses *training* terhadap performa model yang dihasilkan. Performa model akan dihitung berdasarkan *test accuracy* pada setiap *epoch* dan nilai *loss* pada setiap loss (yang direkam setiap 3000 step) pada proses *training*.



Gambar 5 Grafik Relasi Antara Step Dengan Loss



Gambar 6 Grafik Relasi Antara Epoch Dengan Akurasi

Grafik pada Gambar 5 dengan sumbu *x* jumlah *step* dan sumbu *y* nilai *loss* model menunjukkan hubungan antara *step* dengan *loss* dari model. Dari grafik ini dapat dilihat bahwa *step* yang memberikan hasil model paling optimal adalah 12900 *step* atau sama dengan 6 *epoch* dengan tiap *epoch*-nya terdiri dari 2100 *step*. Sementara grafik pada Gambar 6, dengan sumbu *x* jumlah *epoch* dan sumbu *y* nilai akurasi, juga menunjukkan hal yang sama dengan grafik sebelumnya. *Epoch* yang menghasilkan model paling optimal adalah *epoch* ke 6.

### 5.2 Pengujian Akurasi, Presisi, Recall, dan Specifity Model.

Beberapa parameter digunakan untuk menguji performa model. Parameter yang digunakan antara lain adalah akurasi, presisi, recall, dan specifity dari model yang dihasilkan. Proses *training* menggunakan parameter optimal yang dihasilkan dari pengujian sebelumnya pada sub-bab 5.1

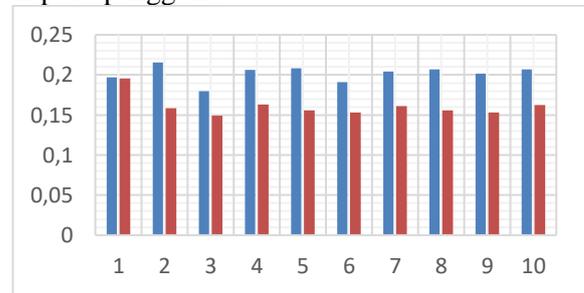
Dari Tabel 1, dengan nilai paling atas akurasi, diikuti dengan presisi, recall, dan specitivity, didapatkan bahwa parameter model yang diuji mendapatkan hasil yang sangat memuaskan. Nilai parameternya berkisar di angka 90% - 100%. Nilai ini dinilai sangat memuaskan untuk model melakukan estimasi arah pandangan pengguna.

Tabel 1 Akurasi, Preisi, Recall, dan Specivity Model

Kelas				
0	1	2	3	4
0.989	0.984	0.993	0.997	0.999
0.947	0.975	0.987	1.000	1.000
1.000	0.950	0.981	0.987	0.993
0.987	0.993	0.997	1.000	1.000

### 5.3 Pengujian Pengaruh CUDA Terhadap Performa

CUDA memiliki pengaruh besar dalam melakukan *inference* terutama pada *single board computer* seperti Jetson TX2 yang memiliki microprocessor yang didesain untuk memakan daya listrik yang kecil. Berdasarkan hasil pengujian yang ditampilkan pada Gambar 7, dengan sumbu *x* adalah pengujian ke- dan sumbu *y* adalah waktu dalam detik, dapat disimpulkan bahwa penggunaan CUDA dapat mempercepat proses *inference* hingga 20% dibandingkan tanpa menggunakan CUDA. Hal ini dinilai sudah sangat baik dalam menjalankan estimasi jarak pandang dan memberikan hasil yang cepat kepada pengguna.



Gambar 7 Perbandingan Pengujian Waktu Komputasi

### 5.4 Pengujian Penggunaan Daya Komputasi

Pada pengujian ini diguankkan rata rata penggunaan CPU dan RAM pada Jetson TX2 saat melakukan *inference* dibandingkan dengan sistem ketika dalam keadaan *idle*. Dari TABEL dapat disimpulkan bahwa penggunaan CPU naik 5 kali lipat dan penggunaan RAM naik hingga 2 kali lipat ketika menjalankan estimasi arah pandangan pengguna. Walaupun kenaikannya yang signifikan namun penggunaan keseluruhan daya komputasi sistem dinilai tidak begitu tinggi sehingga sistem dapat tetap menjalankan *task* lainnya yang perlu dijalankan seperti sistem kendali motor untuk kursi roda pintar dan algoritme *obstacle avoidance* yang lainnya.

Tabel 2 Pengujian Penggunaan Daya Komputasi

Pengujian Penggunaan Daya Komputasi		
System Resource	Idle	Fusion CNN
Overall CPU	5%	26%
RAM	1.4 GiB	3.2 GiB

## 6. KESIMPULAN

Kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut,

1. Dapat disimpulkan bahwa saat meningkatkan epoch, terjadi peningkatan pada akurasi dan penurunan pada hasil dari fungsi loss. Namun, perubahan ini mulai melambat setelah mencapai epoch ke-6, di mana nilai loss tidak mengalami perubahan yang signifikan dan cenderung beresilasi pada nilai yang sama. Akurasi model juga mencapai nilai optimal pada epoch ke-6 dan setelah itu tidak terjadi perubahan yang signifikan. Sehingga, epoch yang optimal untuk mencapai performa terbaik pada model adalah 6.
  2. Hasil pengujian menunjukkan bahwa model dengan parameter training optimal memiliki nilai akurasi rata-rata 0.993 dengan akurasi per kelas terendah sebesar 0.984. Presisi rata-rata model sebesar 0.982 dengan nilai presisi per kelas terendah sebesar 0.947. Nilai recall rata-rata model sebesar 0.982 dengan nilai per kelas terendah sebesar 0.95. Dan nilai specificity rata-rata model sebesar 0.995 dengan nilai per kelas terendah sebesar 0.987
  3. Hasil pengujian menunjukkan bahwa menggunakan akselerasi dengan CUDA dapat menurunkan waktu inference atau waktu komputasi dari algoritme sebesar 20% dibandingkan tanpa menggunakan akselerasi CUDA. Waktu komputasi yang dihasilkan juga memiliki standar deviasi sebesar 0.01, yang menunjukkan bahwa sistem dapat menjalankan algoritme dengan stabil pada waktu 0.16 detik.
  4. Hasil pengujian menunjukkan bahwa penggunaan daya komputasi sistem meningkat sebesar 80% ketika sistem sedang melakukan aktivitas, dibandingkan saat sistem tidak melakukan apa-apa (idle). Namun, rata-rata penggunaan CPU sistem masih tidak mencapai 30%, sehingga sistem masih memiliki kemampuan untuk menjalankan beberapa background task dengan baik. Ada juga kenaikan penggunaan memori sistem hingga 56% atau 3.2GiB dari 4GiB yang tersedia.
- Akinyelu, A. A., & Blignaut, P. (2022). Convolutional Neural Network-Based Technique for Gaze Estimation on Mobile Devices. *Frontiers in Artificial Intelligence*, 4. <https://doi.org/10.3389/frai.2021.796825>
- Anam, K., & Saleh, A. (2020). Voice Controlled Wheelchair for Disabled Patients based on CNN and LSTM; Voice Controlled Wheelchair for Disabled Patients based on CNN and LSTM. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*. <https://doi.org/10.1109/ICICoS51170.2020.9299007>
- Araujo, J. M., Zhang, G., Hansen, J. P. P., & Puthusserypady, S. (2020, June 2). Exploring Eye-Gaze Wheelchair Control. *Eye Tracking Research and Applications Symposium (ETRA)*. <https://doi.org/10.1145/3379157.3388933>
- Fischer, T., Chang, J., & Demiris, Y. (2018). RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments. *European Conference on Computer Vision (ECCV)*, 334–352. [www.imperial.ac.uk/PersonalRobotics](http://www.imperial.ac.uk/PersonalRobotics).
- Mohan, S., & Phirke, M. (2020). Eye Gaze Estimation Invisible and IR Spectrum for Driver Monitoring System. *Signal & Image Processing: An International Journal*, 11(5), 1–20. <https://doi.org/10.5121/sipij.2020.11501>
- Pangestu, G., & Utaminigrum, F. (2020). Electric Wheelchair Control Mechanism Using Eye-mark Key Point Detection. *International Journal of Intelligent Engineering and Systems*, 13(2), 228–238. <https://doi.org/10.22266/ijies2020.0430.22>
- Rockland, R. H., & Reisman, S. (1998). Voice Activated Wheelchair Controller. *Bioengineering, Proceedings of the Northeast Conference*, 128–129. <https://doi.org/10.1109/nebc.1998.664900>
- Yoon, H. S., Baek, N. R., Truong, N. Q., & Park, K. R. (2019). Driver Gaze Detection Based on Deep Residual Networks Using the Combined Single Image of Dual Near-Infrared Cameras. *IEEE Access*, 7,

## 7. DAFTAR PUSTAKA

- Aihara, S., Shibata, R., Mizukami, R., Sakai, T., & Shionoya, A. (2022). Deep Learning-Based Myoelectric Potential Estimation Method for Wheelchair Operation. *Sensors*, 22(4). <https://doi.org/10.3390/s22041615>

93448–93461.  
<https://doi.org/10.1109/ACCESS.2019.2928339>