

Pengenalan Ekspresi Wajah untuk Identifikasi Psikologis Pengguna dengan *Neural Network* dan Transformasi *Ten Crops*

Saeful Bahri¹, Riza Samsinar², Panggalih Sako Denta³

¹⁾²⁾³⁾ Teknik Elektro Universitas Muhammadiyah Jakarta

Jl. Cempaka Putih Tengah 27 No 47

Email: ¹⁾ saefulb@ftumj.ac.id, ²⁾ riza.samsinar@ftumj.ac.id, ³⁾ dentapsako@gmail.com

ABSTRAK

Pada penelitian ini, dirancang sebuah sistem pengenalan ekspresi wajah untuk identifikasi psikologis pengguna dari data mentah tersimpan dengan model neural network yang dirancang berdasarkan AlexNet dan VGG19 dengan 1 dropout layer tanpa ReLU layer pada classifier layer untuk mengurangi jumlah kebutuhan memori yang dipakai pada GPU dan waktu proses secara signifikan sehingga dapat digunakan pada perangkat berdaya komputasi terbatas. Dataset yang digunakan adalah CK+ dataset dan FER2013 dataset dengan transformasi ten crops. Didapatkan waktu proses yang lebih singkat secara signifikan dibandingkan AlexNet dan VGG19 dan tingkat akurasi sebesar 96.970% menggunakan CK+ dataset dan akurasi sebesar 69.128% dan 70.242% menggunakan public test dan private test FER2013 dataset secara berurutan.

Kata Kunci : *deep learning, neural network, data mentah, pengenalan ekspresi, psikologis.*

ABSTRACT

In this research, a facial expression recognition system was designed for psychological identification of users from stored raw data with a neural network model designed based on AlexNet and VGG19 with 1 dropout layer and without a ReLU layer on the classifier layer to significantly reduce the amount of memory used on the GPU and processing time so it can be used on devices with limited computational power. The datasets used are CK+ dataset and FER2013 dataset with ten crops transformation. Obtained significantly shorter processing time than AlexNet and VGG19 and an accuracy rate of 96,970% using the CK + dataset and an accuracy of 69,128% and 70,242% using the public test and private test of FER2013 dataset, respectively.

Keywords : *deep learning, neural network, raw data, expression recognition, psychological.*

1 PENDAHULUAN

Pengenalan ekspresi wajah merupakan kemampuan esensial untuk hubungan interpersonal yang baik, dan merupakan sebuah subyek mayor dalam studi pada bidang perkembangan manusia (*human development*), kesehatan psikologis (*psychological well-being*), dan penyesuaian sosial (*social adjustment*). Faktanya, pengenalan emosi memainkan peranan penting dalam mengalami rasa empati, dalam prediksi perilaku prososial, dan dalam model kemampuan kecerdasan emosional. Kemampuan untuk mengenali emosi dasar muncul sangat dini sejak lahir, sebagaimana kanak-kanak menggunakan ekspresi emosional sebagai penunjuk perilaku [1]

Dalam pengolahan citra digital, penerapan *artificial intelligence* (AI) seperti *deep learning* dengan *convolutional Neural Networks* (CNN) pun belakangan ini meraih kesuksesan dalam pengenalan citra dan video skala besar dan menjadikan CNN sebagai *neural network* yang populer digunakan dan dirancang [2], [3]. Terutama karena ketersediaan

repositori citra publik yang besar untuk emosi dasar [4], seperti CK+ dan FER2013 [5]–[7].

Implementasi GPU optimasi tinggi pada konvolusi 2D dan operasi lainnya dalam pelatihan CNN pada model AlexNet terbukti dapat membantu meningkatkan efisiensi pelatihan model CNN [8]. CUDA Toolkit merupakan salah satu modul yang dapat digunakan dalam pengimplementasian GPU ini [9].

Model VGG pun yang dirancang dengan sampai 19 lapisan *weights* menjelaskan bahwa representasi kedalaman suatu model memengaruhi pada akurasi klasifikasi, dan *state-of-art performance* pada ImageNet *challenge* dataset dan diraih menggunakan arsitektur CNN konvensional dengan menambahkan kedalaman secara substansial [10].

Dengan augmentasi data menggunakan transformasi *ten crops* membuat data input menjadi lebih *robust* [11]. Ditambah dengan tersedianya modul python seperti Pytorch [12], OpenCV [13], Pillow [14], Scikit-image [15], Numpy [16], CSV [17], dan H5PY [18] yang dapat membantu dalam proses pelatihan model, pengolahan citra, dan

pengolahan data. Serta modul *haar cascade* yang dapat membantu untuk mendeteksi wajah [19].

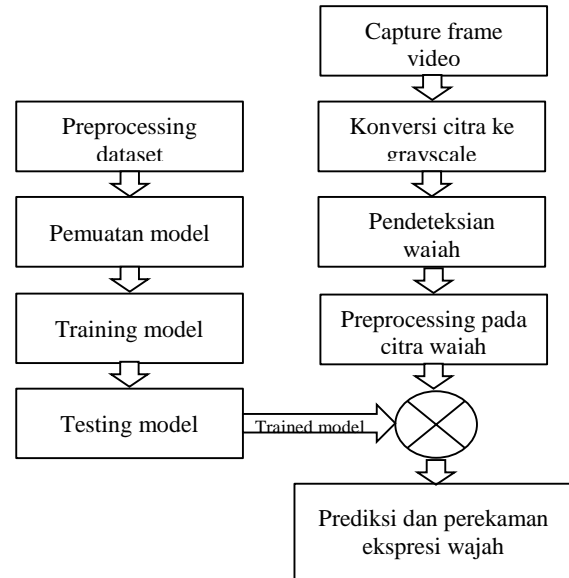
Pengakuisisian dan penyimpanan data prediksi ekspresi menjadi data mentah (*raw data*) merupakan salah satu penerapan FER [20]–[22]. Data tersebut kemudian diolah untuk memantau dan mengidentifikasi psikologis pengguna [23], [24] sebagaimana dokter mendiagnosa pasiennya melalui riwayat penyakit yang pasien miliki.

Akan tetapi model CNN umumnya membutuhkan kekuatan komputasi yang sangat besar, sehingga komputer dengan kekuatan komputasi terbatas akan mengalami kesulitan apabila menggunakan model-model tersebut. Sehingga dibutuhkan CNN yang ringan dan cepat untuk digunakan pada perangkat berdaya komputasi terbatas [25].

Pada penelitian ini, dirancang sistem untuk mengenali ekspresi wajah manusia menggunakan CNN terancang yang ringan dan cepat dalam *deep learning* dengan hanya menggunakan 1 *dropout layer* tanpa ReLU *layer* pada *classifier layer*. Digunakan transformasi *ten crops* untuk input yang lebih *robust*, beberapa modul Python seperti Pytorch untuk proses *deep learning*; OpenCV, Pillow, dan Scikit-iaage untuk pengolahan citra; Numpy, CSV, dan H5PY untuk mengolah data, serta menggunakan 1 CUDA-enabled GPU driver NVIDIA GeForce MX150 dengan CUDA Toolkit untuk penggunaan GPU, dan *haar cascade* sebagai detector wajah. Lalu diaplikasikan pada merekam prediksi ekspresi wajah menjadi data mentah, data tersebut kemudian digunakan untuk memantau dan menunjukkan riwayat ekspresi wajah pengguna terprediksi untuk mengidentifikasi psikologis pengguna. Dataset yang digunakan adalah CK+ dan FER2013.

2 METODOLOGI

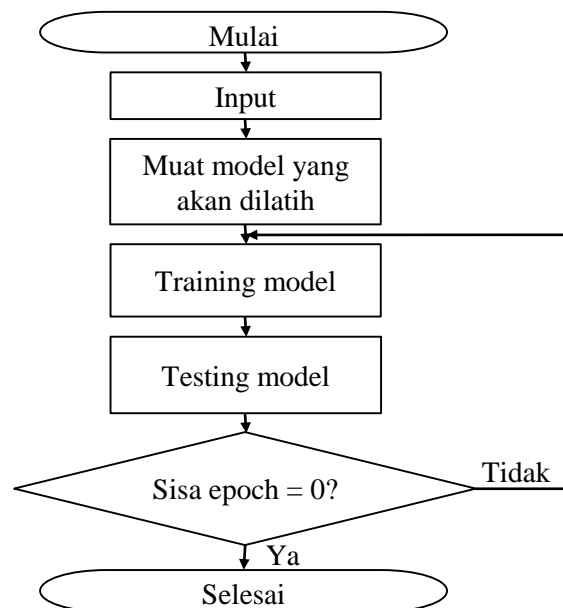
Pada penelitian ini, diterapkan metode *deep learning* dan dirancang model *convolutional neural network* yang mengacu pada AlexNet dan VGG terkostumisasi dengan 1 *dropout layer* tanpa ReLU *layer* pada *classifier layer*. Digunakan transformasi *ten crops*, beberapa modul Python yang telah disebutkan, 1 CUDA-enabled GPU driver NVIDIA GeForce MX150 dengan CUDA Toolkit, dan detector wajah *haar cascade*, serta CK+ dan FER2013 sebagai datasetnya. Lalu dari hasil prediksi ekspresi wajah menggunakan model yang dirancang, data akan disimpan menjadi *raw data* untuk digunakan pada identifikasi psikologis pengguna. Kemudian sistem akan dievaluasi berdasarkan tingkat akurasi prediksi antara label input dan outputnya dan berdasarkan lama waktu proses dibanding dengan model AlexNet dan VGG.



Gambar 1 Diagram kerja sistem secara umum.

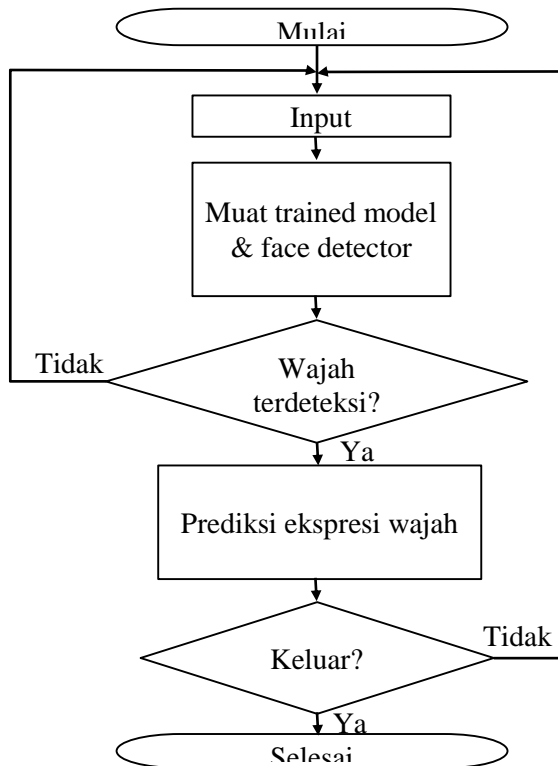
Secara umum, alur kerja sistem terbagi menjadi dua. Yaitu alur pertama untuk perancangan dan pelatihan model menjadi *trained model* dan alur kedua untuk simulasi prediksi ekspresi wajah dan perekaman data prediksi menjadi data mentah yang dapat dilihat pada Gambar 1. Citra terakuisisi diubah menjadi citra abu-abu (*grayscale*) dengan ukuran 48x48, dan mentransformasi citra menggunakan transformasi *ten crops* dengan ukuran potong sebesar 44x44 seperti pada proses pengujian model.

Gambar 2 menunjukkan *flowchart* pada proses pelatihan model,



Gambar 2 Flowchart proses pelatihan model.

Gambar 3 berikut menunjukkan *flowchart* sistem pada saat sistem disimulasikan,



Gambar 3 *Flowchart* sistem.

Dalam pelaksanaan penelitian ini, dibutuhkan beberapa perangkat keras (*hardware*) dan perangkat lunak (*software*) yang ditunjukkan pada tabel 1 dan tabel 2 sebagai berikut,

Tabel 1 Kebutuhan *hardware*.

No.	Nama	Fungsi
1.	Komputer atau Laptop	Perangkat pengerjaan
2.	CUDA-Enabled NVIDIA GPU	GPU yang digunakan dalam pelatihan
3.	Webcam	Input citra dalam pengujian

Tabel 2 Kebutuhan *software*.

No.	Nama	Fungsi
1.	Python	Bahasa Pemrograman
2.	CUDA Toolkit	Toolkit pemrograman dengan GPU

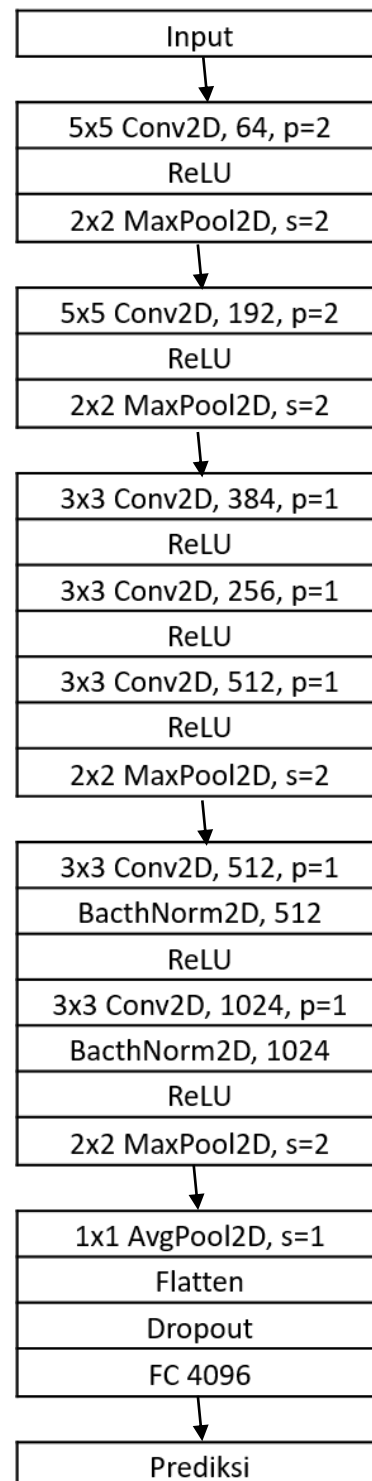
Adapun spesifikasi minimum *hardware* yang dibutuhkan antara lain :

CPU : Intel Core i5-8250U @ 1.60GHz (2 CPUs)
 GPU : NVIDIA GeForce MX150

RAM : 2GB

ROM : 5GB Free Spaces

Model yang dirancang berupa penggabungan antara lapisan konvolusi pada model AlexNet yang termodifikasi dengan lapisan akhir konvolusi VGG dengan *batch normalization* yang telah termodifikasi untuk lapisan konvolusi model yang dirancang yang kemudian dihubungkan dengan 1 lapisan *fully-connected layer* (FCL).



Gambar 4 Arsitektur model yang akan dirancang.

Arsitektur model *neural network* yang dirancang pada penelitian ini berbentuk seperti pada gambar 5, model yang dirancang memiliki total 7 lapisan Conv2D disertai dengan fungsi ReLU, 4 lapisan MaxPool2D dengan 2x2 kernel dan stride sebesar 2, dan 2 lapisan BatchNorm2D pada 2 lapisan akhir konvolusi. Pada 2 lapisan konvolusi pertama digunakan 5x5 kernel dengan padding sebesar 2, lalu pada 5 lapisan konvolusi berikutnya digunakan 3x3 kernel dengan padding sebesar 1. Lalu dihubungkan dengan lapisan AvgPool2D dengan 1x1 kernel dan stride sebesar 1, digunakan fungsi flatten dan dropout sehingga dengan *fully-connected layer* sebesar 4096.

Parameter yang diperoleh dari model *neural network* yang dirancang dapat dilihat pada gambar 5 berikut,

Layer (type)	Output Shape	Param #
Conv2d-1	[-1, 64, 44, 44]	4,864
ReLU-2	[-1, 64, 44, 44]	0
MaxPool2d-3	[-1, 64, 22, 22]	0
Conv2d-4	[-1, 192, 22, 22]	307,392
ReLU-5	[-1, 192, 22, 22]	0
MaxPool2d-6	[-1, 192, 11, 11]	0
Conv2d-7	[-1, 384, 11, 11]	663,936
ReLU-8	[-1, 384, 11, 11]	0
Conv2d-9	[-1, 256, 11, 11]	884,992
ReLU-10	[-1, 256, 11, 11]	0
Conv2d-11	[-1, 512, 11, 11]	1,180,160
ReLU-12	[-1, 512, 11, 11]	0
MaxPool2d-13	[-1, 512, 5, 5]	0
Conv2d-14	[-1, 512, 5, 5]	2,359,808
BatchNorm2d-15	[-1, 512, 5, 5]	1,024
ReLU-16	[-1, 512, 5, 5]	0
Conv2d-17	[-1, 1024, 5, 5]	4,719,616
BatchNorm2d-18	[-1, 1024, 5, 5]	2,048
ReLU-19	[-1, 1024, 5, 5]	0
MaxPool2d-20	[-1, 1024, 2, 2]	0
AvgPool2d-21	[-1, 1024, 2, 2]	0
Linear-22	[-1, 7]	28,679

Total params: 10,152,519
 Trainable params: 10,152,519
 Non-trainable params: 0

Gambar 5 Parameter model *neural network* yang dirancang.

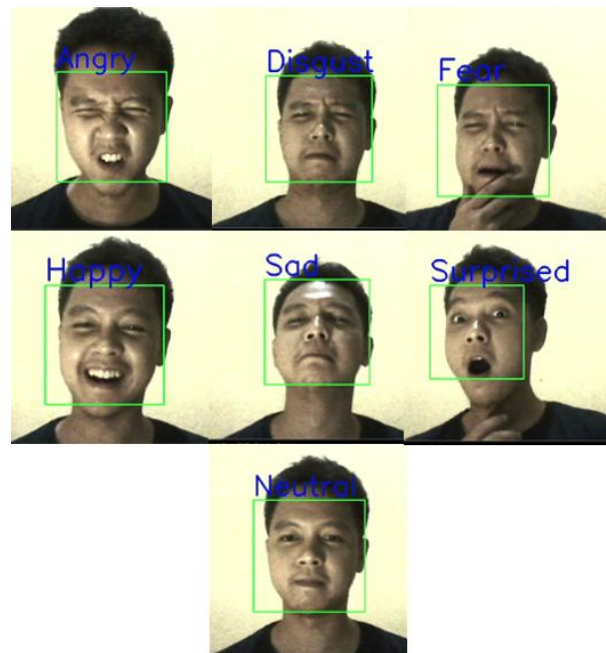
Adapun ekspresi wajah yang digunakan diklasifikasikan sebagai berikut :

1. *Angry* dengan label angka 0
2. *Disgust* dengan label angka 1
3. *Fear* dengan label angka 2
4. *Happy* dengan label angka 3
5. *Sad* dengan label angka 4
6. *Surprise* dengan label angka 5
7. *Neutral* dengan label angka 6

3 HASIL DAN PEMBAHASAN

Proses perhitungan prediksi ekspresi wajah menggunakan fungsi softmax untuk menghitung nilai loss dan mencari nilai max pada nilai prediksi yang diperoleh. Nilai max ini yang digunakan untuk

menentukan ekspresi wajah yang diprediksi. Hasil prediksi ekspresi wajah pada saat simulasi dapat dilihat pada Gambar 6,



Gambar 6 Simulasi prediksi ekspresi wajah.

Kemudian data yang diperoleh dari hasil prediksi ekspresi wajah akan disimpan menjadi data mentah (*raw data*) dengan bentuk seperti pada Gambar 8,

23:41.2	3-Aug-20	[1.7103852	4 Sad
23:41.5	3-Aug-20	[3.3786182	4 Sad
23:41.5	3-Aug-20	[3.3786182	4 Sad
23:41.7	3-Aug-20	[5.1495568	4 Sad

Gambar 8 Sampel data mentah yang tersimpan.

Hasil evaluasi berdasarkan pada waktu proses sebanyak 5 *epoch* ditunjukkan pada Tabel 3 dan Tabel 4 serta hasil evaluasi berdasarkan pada akurasi prediksi ekspresi wajah dengan model yang dirancang dapat dilihat pada Tabel 5 sampai Tabel 7.

Tabel 3 Evaluasi waktu dengan CK+ dataset.

Model	Waktu (s)		Waktu Total (s)
	Train	Test	
VGG19	28.43	12.38	40.81
AlexNet	26.43	17.49	43.92
Model yang dirancang	17.73	7.85	25.58

Tabel 4 Evaluasi waktu dengan FER2013 dataset.

Model	Waktu (s)			Waktu Total (s)
	Train	Public Test	Private Test	
VGG19	2741.87	419.57	423.96	3585.4
AlexNet	1458.25	687.8	679.08	2825.13
Model yang dirancang	1655.82	256.6	257.34	2169.76

Tabel 5 Evaluasi akurasi dengan CK+ dataset.

Label Ekspresi	Jumlah	Benar	Salah	Akurasi (%)
Angry	12	12	0	100
Disgust	18	18	0	100
Fear	9	3	0	100
Happy	21	21	0	100
Sad	9	6	3	66.667
Surprise	24	24	0	100
Contempt	6	6	0	100
Total akurasi				96.970

Tabel 6 Evaluasi akurasi dengan *public test* FER2013 dataset.

Label Ekspresi	Jumlah	Benar	Salah	Akurasi (%)
Angry	467	299	168	64.026
Disgust	56	38	18	67.857
Fear	496	228	268	45.968
Happy	895	778	117	86.927
Sad	653	405	248	62.021
Surprise	415	357	58	86.024
Neutral	607	376	231	61.944
Total akurasi				69.128

Tabel 7 Evaluasi akurasi dengan *private test* FER2013 dataset.

Label Ekspresi	Jumlah	Benar	Salah	Akurasi (%)
Angry	491	300	191	61.100
Disgust	55	39	16	70.909
Fear	528	248	280	46.970
Happy	879	783	96	89.078
Sad	594	357	237	60.101
Surprise	416	362	54	87.019
Neutral	626	432	194	69.010
Total akurasi				70.242

4 KESIMPULAN

Kesimpulan yang didapat dari penelitian ini adalah sebagai berikut :

1. Kostumisasi model dengan mengurangi dropout layer menjadi 1 pada classifier layer dapat mengurangi waktu proses.
2. Diketahui bahwa model yang dirancang membutuhkan waktu proses yang jauh lebih singkat dibanding AlexNet dan VGG dengan waktu keseluruhan proses sebanyak 5 epoch yaitu 25.58 detik untuk CK+ dataset dan 2169.76 detik untuk FER2013 dataset.
3. Tingkat akurasi menggunakan CK+ dataset lebih tinggi yaitu sebesar 96.970% dibandingkan dengan menggunakan FER2013 dataset yaitu sebesar 69.128% dan 70.242%, karena CK+ dataset terkompilasi dari hasil foto di laboratorium dengan, sudut pandang yang sama, pose yang relatif sama, dan jumlah data yang jauh lebih sedikit sehingga dataset mudah sekali mengakibatkan bias.

DAFTAR PUSTAKA

- [1] G. Mancini, R. Biolcati, S. Agnoli, F. Andrei, and E. Trombini, "Recognition of facial emotional expressions among Italian pre-adolescents, and their affective reactions," *Front. Psychol.*, 2018, doi: 10.3389/fpsyg.2018.01303.
- [2] S. Samoili, M. López Cobo, G. Gómez, E., De Prato, F. Martínez-Plumed, and B. Delipetrev, *AI Watch - Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence.* 2020. doi: 10.2760/382730.
- [3] F. Chollet, *Deep Learning with Phyton.* 2018.
- [4] P. Ekman, "Basic Emotions," in *Handbook of Cognition and Emotion*, 2005. doi: 10.1002/0470013494.ch3.
- [5] P. Lucey *et al.*, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," no. July, 2010.
- [6] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," 2000. doi: 10.1109/AFGR.2000.840611.
- [7] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Netw.*, 2015, doi: 10.1016/j.neunet.2014.09.005.

- [8] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," 2014.
- [9] NVIDIA Corporation, "About CUDA," in <https://developer.nvidia.com/about-cuda>, 2015.
- [10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
- [11] Z. Qin, J. Wu, Y. Liu, and T. Gedeon, "Visual Saliency Maps Can Apply to Facial Expression Recognition," pp. 1–8.
- [12] A. Paszke *et al.*, "Automatic differentiation in PyTorch," 2019.
- [13] G. Bradski, "The OpenCV Library," *Dr Dobbs J. Softw. Tools*, 2000, doi: 10.1111/0023-8333.50.s1.10.
- [14] Y. Mori and T. Igarashi, "Pillow," 2006. doi: 10.1145/1179849.1179942.
- [15] S. Van Der Walt *et al.*, "Scikit-image: Image processing in python," *PeerJ*, 2014, doi: 10.7717/peerj.453.
- [16] T. Oliphant and J. k. Millma, "A guide to NumPy," *Trelgol Publishing*. 2006. doi: DOI:10.1109/MCSE.2007.58.
- [17] G. Van Rossum and F. L. Drake, "The Python Library Reference," *October*, 2010.
- [18] C. M. Lewandowski, N. Co-investigator, and C. M. Lewandowski, *Python and HDF5*. 2015. doi: 10.1017/CBO9781107415324.004.
- [19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," 2001. doi: 10.1109/cvpr.2001.990517.
- [20] I. Isnaini, P. S. Denta, R. S. Hakiki, M. Amin, and J. Pusat, "PEMBUATAN PROGRAM UNTUK MENENTUKAN PUNCAK SPEKTRUM," vol. 16, pp. 33–41, 2019.
- [21] S. Bahri and H. Kusindaryadi, "Rancang Bangun Pemantauan Absensi Mahasiswa dengan Menggunakan Sidik Wajah secara Simultan Melalui CCTV Ruang Kelas," *Resist. Elektron. KEndali Telekomun. Tenaga List. Komput.*, vol. 3, no. 1, p. 37, 2020, doi: 10.24853/resistor.3.1.37-44.
- [22] R. Samsinar, R. Septian, and F. Fadliondi, "Alat Monitoring Suhu Kelembapan dan Kecepatan Angin dengan Akuisisi Database Berbasis Raspberry Pi," *Resist. Elektron. KEndali Telekomun. Tenaga List. Komput.*, vol. 3, no. 1, p. 29, 2020, doi: 10.24853/resistor.3.1.29-36.
- [23] H. Isyanto and A. Nandiwardhana, "Perancangan DC Cooler Berbasis Internet of Things," *Resist. Elektron. KEndali Telekomun. Tenaga List. Komput.*, vol. 2, no. 2, p. 95, 2019, doi: 10.24853/resistor.2.2.95-104.
- [24] H. Muchtar and F. Said, "Sistem Identifikasi Plat Nomor Kendaraan Menggunakan Metode Robert Filter dan Framing Image Berbasis Pengolahan Citra Digital," *Resist. Elektron. KEndali Telekomun. Tenaga List. Komput.*, vol. 2, no. 2, p. 105, 2019, doi: 10.24853/resistor.2.2.105-112.
- [25] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The Computational Limits of Deep Learning," 2020.