

Custom Convolutional Neural Network (CNN) Model untuk Pengenalan Emosi pada Real-Time Video

Esa Sulistyo Aji Nugroho^{a1}, I Wayan Santiyasa^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹nugroho.2208561144@student.unud.ac.id
²santiyasa@unud.ac.id

Abstract

Facial Expressions Recognition (FER) is a vital task in human-computer interaction. This field can be used to augment many other fields such as learning, content suggestions and many more. This paper aims to propose a simple FER approach utilizing a custom Convolutional Neural Network (CNN) architecture for real-time emotion recognition, complemented by Hard Cascades for efficient face detection. The model is trained on a large and reputable dataset encompassing various facial expressions classified into 7 main emotions. The result of this study yields an accuracy of 60.78% achieved on the training data and 57.32% on the validation data. This shows that with even a very simplified approach with a custom CNN model, we can relatively accurately recognize facial expressions. This study aims to be a kind of groundwork for more advanced approaches to FER.

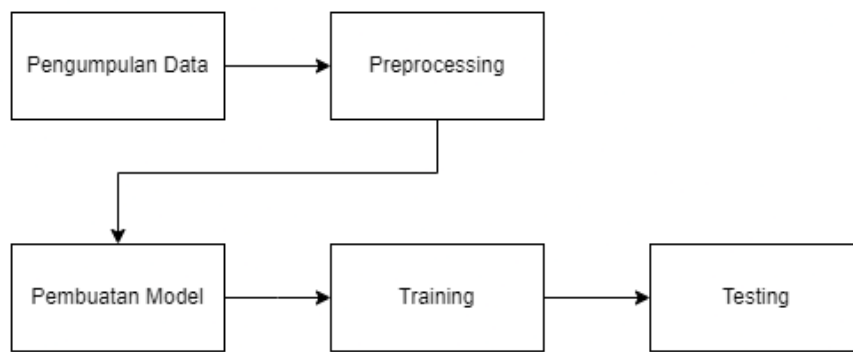
Keywords: Image Processing, Convolutional Neural Network, Deep Learning, Image Classification, Facial Emotion Recognition

1. Pendahuluan

Ekspresi wajah merupakan aspek mendasar dari komunikasi manusia yang menyampaikan emosi, niat, dan reaksi [1]. Pengenalan emosi otomatis adalah bidang penelitian besar dan penting yang membahas dua subjek berbeda, yaitu pengenalan emosi psikologis manusia dan kecerdasan buatan (AI) [2]. Pengenalan emosi wajah secara *real-time* menawarkan pintu gerbang untuk memahami ekspresi manusia dengan kedalaman dan ketepatan yang belum pernah terjadi sebelumnya. Pengenalan wajah dimulai pada tahun 1960-an ketika tim peneliti yang dipimpin oleh W. Bledsoe melakukan eksperimen untuk menentukan apakah komputer dapat mengenali wajah manusia. Tim Bledsoe berusaha membangun hubungan antara detail wajah manusia sehingga komputer dapat menemukan serangkaian kecocokan yang memungkinkan pengenalan wajah tersebut [3][4]. Dalam beberapa tahun terakhir, terdapat peningkatan minat dalam menggunakan teknik *Machine Learning* untuk mendeteksi dan mengklasifikasikan emosi secara otomatis. Salah satu teknik canggih tersebut adalah *Convolutional Neural Network* (CNN) [5]. Artikel ini mengeksplorasi perpaduan *Convolutional Neural Network* (CNNs) dan *Hard Cascades*, dua teknik ampuh yang telah banyak digunakan dalam berbagai aplikasi pengenalan emosi.

2. Metode Penelitian

Proses penelitian ini akan dilakukan seperti di Gambar 1.



Gambar 1. *Timeline Penelitian*

2.1. Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari *dataset* FER2013. *Dataset* tersebut merupakan *dataset* untuk *benchmark* model-model FER yang terdiri dari hampir 35.887 gambar *grayscale* 8-bit berukuran 48×48 -pixel dari berbagai ekspresi wajah orang. *Dataset* tersebut dibagi menjadi 28,000 data *training* dan 3,500 data validasi.

Untuk Preprocessing hanya perlu dilakukan pada saat menginisialisasi data untuk *training* dan validasi karena *Dataset* FER2013 sudah terstruktur dan rapi.

2.2. Convolutional Neural Network (CNN)

CNN dibentuk oleh beberapa lapisan *neuron*, dengan lapisan konvolusional menjadi yang paling penting. Pengoperasiannya didasarkan pada penerapan serangkaian filter yang bergerak melalui gambar untuk mendapatkan keluaran lapisan [6].

Dalam penelitian ini algoritma yang digunakan untuk mendeteksi emosi wajah merupakan CNN dan untuk mendeteksi wajahnya sendiri dalam video merupakan algoritma *Hard Cascades* tepatnya *Hard Cascade Frontalface* yang mendeteksi wajah dari depan.

Berikut merupakan desain dari model CNN yang akan digunakan.

Tabel 1. Struktur Model CNN

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 32)	544
max_pooling2d (MaxPooling2D)	(None, 24, 24, 32)	0
conv2d_1 (Conv2D)	(None, 22, 22, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 11, 11, 64)	0
dropout_1 (Dropout)	(None, 11, 11, 64)	0
conv2d_2 (Conv2D)	(None, 10, 10, 128)	32,896
max_pooling2d_2 (MaxPooling2D)	(None, 5, 5, 128)	0
conv2d_3 (Conv2D)	(None, 4, 4, 128)	65,664
max_pooling2d_3 (MaxPooling2D)	(None, 2, 2, 128)	0
dropout_1 (Dropout)	(None, 2, 2, 128)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 1024)	525,312

Layer (type)	Output Shape	Param #
dropout_2 (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 256)	262,400
dropout_3 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 256)	65,792
dropout_4 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 7)	1,799

Model tersebut memiliki total *parameter* sebanyak 2,918,711, yang terdiri dari 972,903 *Trainable params* dan 1,945,808 *Optimizer params*.

3. Hasil dan Pembahasan

Pada Bagian ini akan dijabarkan data hasil *training* dan validasi yang meliputi akurasi, *loss*, presisi, *recall* dan *F1-Score* serta demonstrasi model digunakan dengan *real-time camera feed*. *Training* dilakukan dengan menggunakan *batch_size* 64 dan 40 *epoch*.

3.1. Hasil *Training* dan Validasi

Model dibuat untuk mengklasifikasi ekspresi wajah menjadi 7 yaitu marah, jijik, takut, bahagia, netral, sedih dan terkejut. Pada fase *training* didapatkan akurasi 60.74% dengan *loss* 5.47% dan akurasi validasi 58.32% dengan *loss* validasi 9.28%.

Dibawah merupakan tabel hasil validasi performa dari model.

Tabel 2. *Precision, Recall dan F1-score*

Ekspresi	Precision	Recall	F1-Score	Support
0 (marah)	0.14	0.12	0.13	958
1 (jijik)	0.02	0.01	0.01	111
2 (takut)	0.15	0.06	0.01	1024
3 (bahagia)	0.24	0.26	0.25	1774
4 (netral)	0.18	0.23	0.20	1233
5 (sedih)	0.17	0.22	0.19	1247
6 (terkejut)	0.11	0.11	0.11	831

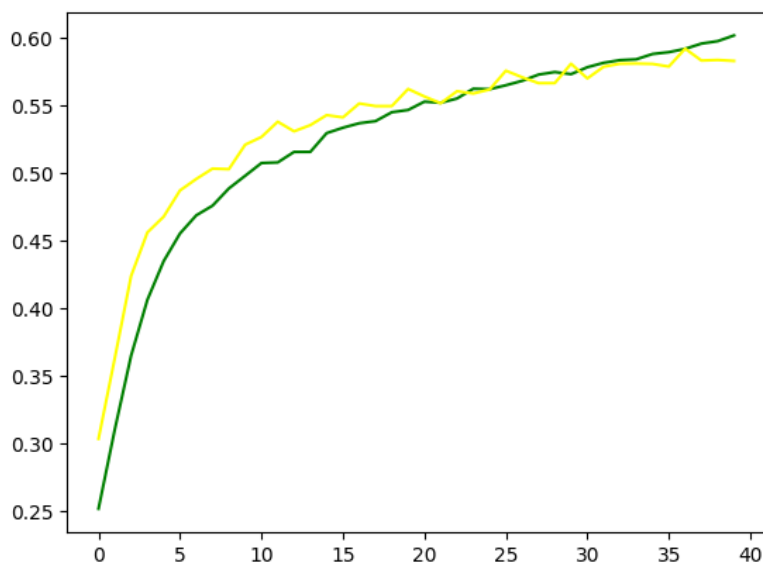
Dari Tabel diatas dapat dilihat presisi, recall dan F1-Score saat model memprediksi masing-masing ekspresi. Dapat juga dilihat bahwa bahagia adalah emosi yang paling konsisten diprediksi.

Adapun graph dan tabel akurasi, akurasi validasi, *loss* dan *loss* validasi saat *training*.

Tabel 3. Akurasi, Akurasi Validasi, *Loss* dan *Loss* Validasi

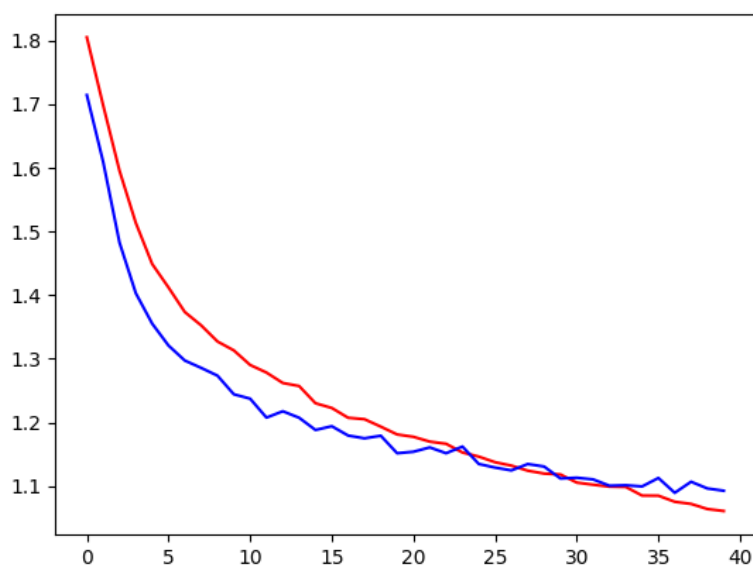
epoch	akurasi (higher is better)	loss (lower is better)	akurasi validasi	validasi loss
0	~25%	~80%	~30%	~70%
10	~50%	~30%	~53%	~25%

epoch	akurasi (higher is better)	loss (lower is better)	akurasi validasi	validasi loss
20	~55%	~20	~55%	~17%
30	~57%	~13%	~57%	~13%
40	60.74%	5.47%	58.32%	9.28%



Gambar 2. *accuracy-val_accuracy*

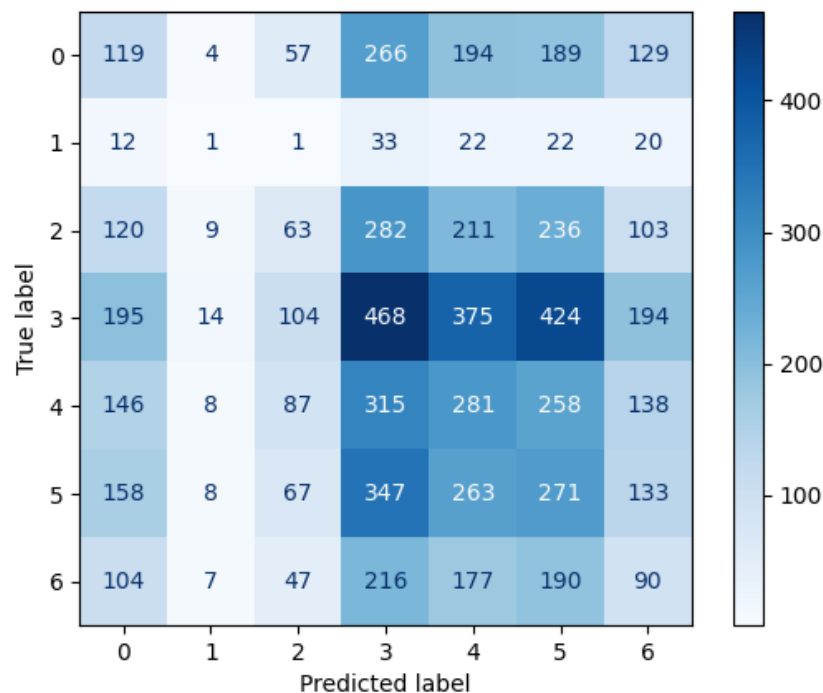
Garis kuning pada gambar 2 merupakan akurasi validasi dan garis hijau adalah akurasi saat *training*, *axis* mendatar(x) merupakan jumlah *epoch training* dan *axis* vertikal(y) merupakan akurasi. Dapat dilihat bahwa akurasi keduanya meningkat dari awal ~25% hingga akhir *training* ~60%.



Gambar 3. *loss-val_loss*

Garis biru pada gambar 3 merupakan *loss* validasi dan garis merah adalah *loss* saat *training*, *axis* mendatar(x) merupakan jumlah *epoch training* dan *axis* vertikal(y) merupakan *loss*. Dapat dilihat bahwa akurasi keduanya menurun dari awal ~80% hingga akhir *training* <10%.

Berikut juga merupakan *Confusion Matrix*.

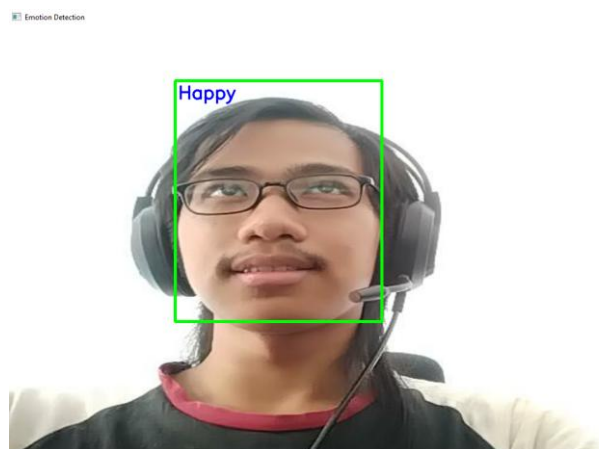


Gambar 4. *Confusion Matrix*

Pada confusion matrix diatas dapat dilihat presisi dari semua ekspresi dimana 0: "marah", 1: "jijik", 2: "takut", 3: "bahagia", 4: "Netral", 5: "sedih" dan 6: "terkejut". bisa dilihat seperti pada tabel 2 ekspresi "bahagia" merupakan ekspresi yang paling sering benar diprediksi dan "jijik" adalah yang paling jarang diprediksi dengan benar.

3.2. Tes Dengan *Real-Time Feed*

Saat Model dicoba dengan *real-time feed* dari kamera terlihat bahwa algoritma mampu mendeteksi wajah dan juga mengenali ekspresi. Terlihat juga bahwa hasil validasi presisi yang ada pada tabel 2 benar karena emosi yang paling sering terdeteksi adalah 'bahagia', 'netral' dan 'takut'.



Gambar 5. *Real-Time Video Feed*

4. Kesimpulan

Berdasarkan penelitian ini ditemukan bahwa arsitektur *custom Convolutional Neural Network* dapat digunakan dalam pemrosesan *Facial Emotion Recognition* (FER) yakni dalam *real-time recognition*. Dari hasil latih model didapatkan akurasi 60.74% untuk data latih dan 58.32% untuk data validasi.

Daftar Pustaka

- [1] R. Lawpanom, W. Songpan and J. Kaewyotha, "Advancing Facial Expression Recognition in Online Learning Education Using a Homogeneous Ensemble Convolutional Neural Network Approach" *Applied Sciences*, vol. 14, no. 3: 1156, 2024.
- [2] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights" *Procedia Computer Science*, vol. 175, p. 689-694, 2020.
- [3] R. Plutchik, "The nature of emotions" *Am. Scientist*, vol. 89, p. 334–350, 2001.
- [4] J. A. Ballesteros, V. Ramírez, M. Gabriel, F. Moreira, A. Solano and C. A. Pelaez, "Facial emotion recognition through artificial intelligence" *Frontiers in Computer Science*, vol. 6, 2024.
- [5] S. Kusawa, "datamagiclab", 15 Juni 2023 [Online], Available: <https://datamagiclab.com/emotion-detection-using-convolutional-neural-network/>. [3 Mei 2024]
- [6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition" in ICLR, San Diego, 2015.