

# Lightweight Convolutional Neural Network Architecture For Facial Expression Recognition

Arsitektur Convolutional Neural Network Ringan Untuk Pengenalan Ekspresi Wajah

Reynold Frankie Robot, Miesje Karenia Lolowang, Muhamad Dwisnanto Putro  
Dept. of Electrical Engineering, Sam Ratulangi University Manado, Kampus Bahu St., 95115, Indonesia  
e-mails : [reynold.robot@unsrat.ac.id](mailto:reynold.robot@unsrat.ac.id), [karen.audrey148@gmail.com](mailto:karen.audrey148@gmail.com), [dwisnantoputro@unsrat.ac.id](mailto:dwisnantoputro@unsrat.ac.id)

Received: 03 February 2024; revised: 25 March 2024 ; accepted: 15 April 2024

**Abstract** — Facial expression recognition presents the challenge of applicability to real scenarios which demands to involve lightweight algorithms to operate at real-time speed. The proposed work offers a lightweight convolutional neural network architecture to accurately predict facial expressions. It considers the use of parameters and computations that are not massive. thus applying lightweight convolution operations. A channel compression technique at the center is applied to reduce parameter usage and redundant operations. Testing the accuracy results was conducted on the KDEF dataset. This dataset is a knowledge source that provides information on seven facial expressions with varying poses. Augmentation techniques were used to increase the variety and training configuration strategies were applied to optimize the network training work. As a result, the proposed model achieves an accuracy of 0.9732 which outperforms competing CNN architectures. Besides, it also produces a lightweight parameter of 2.8 M which can encourage the model to operate fast on Jetson Nano devices. The proposed architecture produces an effective deep learning model for predicting facial expressions from a chunk of input image without compromising its efficiency.

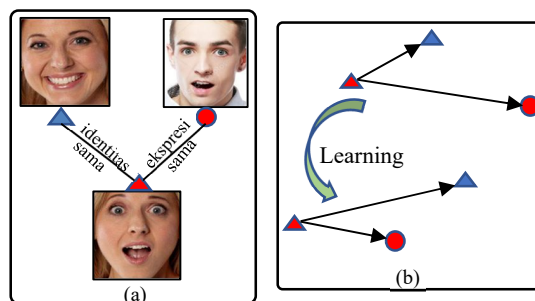
**Key words**— Facial expressions, lightweight architecture, CNN, model efficiency.

**Abstrak** — Pengenalan ekspresi wajah menghadirkan tantangan penerapan pada skenario nyata yang menuntut untuk melibatkan algoritma yang ringan agar dapat beroperasi dengan kecepatan waktu nyata. Pekerjaan yang diusulkan menawarkan sebuah arsitektur *convolutional neural network* yang ringan untuk memprediksi ekspresi wajah secara akurat. Ini mempertimbangan penggunaan parameter dan komputasi yang tidak massif sehingga menerapkan operasi konvolusi yang ringan. Teknik pemampatan kanal pada bagian tengah diterapkan untuk mengurangi penggunaan parameter dan operasi berlebihan. Pengujian hasil akurasi diadakan pada dataset KDEF. Dataset ini merupakan sumber pengetahuan yang menyediakan informasi tujuh ekspresi wajah dengan pose yang bervariasi. Teknik augmentasi digunakan untuk memperbanyak variasi dan strategi konfigurasi pelatihan diterapkan untuk mengoptimalkan pekerjaan pelatihan jaringan. Sebagai hasil, model yang diusulkan meraih akurasi sebesar 0.9732 yang mengungguli arsitektur CNN pesaing. Disamping itu, ini juga menghasilkan parameter yang ringan sebesar 2.8 M yang dapat mendorong model untuk beroperasi cepat pada perangkat Jetson Nano. Arsitektur yang diusulkan menghasilkan model pembelajaran mendalam yang efektif untuk memprediksi ekspresi wajah dari sebuah potongan masukan citra tanpa mengabaikan efisiensinya.  
**Kata kunci** — Ekspresi wajah, arsitektur ringan, CNN, efisiensi model.

## I. PENDAHULUAN

Salah satu teknologi komputer vision yang kini berkembang pesat adalah *facial expression recognition (FER)* dan teknologi FER tidak bisa terlepas dari proses identifikasi emosi dari suatu wajah [1]. Wajah menyampaikan banyak informasi tentang maksud dan perasaan seseorang karena wajah merupakan tempat utama ekspresi, pengenalan, dan komunikasi. Sebagai sarana komunikasi universal, ekspresi wajah memungkinkan orang menyampaikan emosi mereka dengan cara yang dapat dipahami orang lain [2]. Wajah juga merupakan salah satu cara paling ekspresif, universal, dan alami bagi manusia untuk mengkomunikasikan keadaan emosi mereka. Itu sebabnya betapa pentingnya FER otomatis dalam berbagai aplikasi, seperti interaksi manusia-mesin dan perawatan kesehatan [3], sehingga banyak penelitian telah dilakukan tentang hal ini. Namun, karena identitas wajah, posisi wajah, dan pencahayaan yang berbeda dari tiap gambar wajah, sehingga memerlukan upaya yang harus dilakukan untuk meningkatkan akurasi FER [4], [5].

Sebagai contoh yang dapat dilihat dalam Gambar 1, fitur gambar lebih didominasi oleh variasi identitas daripada ekspresi wajah. Akibatnya, jarak fitur antara subjek yang berbeda dengan *ekspresi yang sama* mungkin lebih besar daripada jarak antara *ekspresi yang berbeda* dari subjek yang sama, yang menyiratkan bahwa klasifikasi ekspresi wajah tidak kuat terhadap fitur gambar. Oleh sebab itu, untuk meningkatkan FER, representasi fitur ekspresi yang efisien sangatlah penting [6].



Gambar 1. Gambar wajah berbeda ekspresi (a) dan gambar yang menunjukkan perbedaan jarak fitur (b)

Emosi setiap orang merupakan komponen penting dalam komunikasi mereka. Ada banyak cara emosi dapat disampaikan, dan terkadang sulit untuk mengidentifikasinya dengan jelas [7]. Banyak elemen termasuk ucapan, teks, bahasa tubuh, dan ekspresi wajah, dapat digunakan untuk mengidentifikasi emosi. Hal tersebut dikarenakan ekspresi wajah dapat langsung diamati dengan identifikasi emosi melalui ekspresi wajah yang merupakan salah satu elemen yang paling sering digunakan [8]. Emosi dan kognisi dipandang sebagai "alat sederhana namun bermanfaat" yang membantu seseorang berkembang dan menjalani kehidupan [9]. Melalui ekspresi mikro di media sosial, emosi sebagai respons terhadap rangsangan atau peristiwa tertentu dapat terbaca. Tujuh emosi wajah yang paling umum ditentukan oleh psikolog Ekman, yaitu: senang, sedih, jijik, geram, heran, dan netral. Kategorisasi ini hendak memberikan pemahaman yang mendasar tentang bagaimana memahami emosi seseorang dari ekspresi wajah mereka [2].

Studi ilmiah tentang identifikasi ekspresi wajah sangat menarik, terutama mengingat perkembangan terkini dalam pembelajaran mesin dan kecerdasan buatan. Salah satu teknologi pembelajaran mendalam yang sangat baik dalam mengidentifikasi pola dalam data gambar dengan menggunakan teknik *convolutional neural network* (CNN) [2]. FER telah menjadi hal yang krusial di banyak sektor karena kemajuan teknologi. Keamanan gedung merupakan salah satu area yang membutuhkan FER. Pengenalan wajah telah mulai digunakan dalam sistem pemantauan keamanan gedung untuk mengidentifikasi individu. Kamera dengan kemampuan pengenalan wajah dapat dipasang di ruangan sehingga dapat mengidentifikasi wajah orang yang memasuki atau meninggalkan gedung. FER dapat dilakukan dengan menggunakan data yang diproses ulang yang diperoleh dari pengenalan wajah. Sejumlah informasi yang berguna dihasilkan dengan menggabungkan hasil pengenalan wajah dan pengenalan ekspresi wajah yang dapat dimanfaatkan untuk deteksi, presensi kehadiran, dan pemantauan (*monitoring*). Sayangnya, variasi ekspresi wajah antar subjek dapat menyebabkan penurunan akurasi FER. Metode augmentasi dapat digunakan pada sekumpulan ekspresi wajah untuk meningkatkan akurasi FER. Teknik untuk memodifikasi data tanpa kehilangan esensi atau intinya dikenal sebagai teknik augmentasi. Hasil dari berbagai modifikasi ini akan berfungsi sebagai kumpulan data untuk pengembangan model berikut.

Kumpulan data yang dibuat oleh prosedur augmentasi dapat digunakan dalam metode *transfer learning* sebagai data latih. Metode atau pendekatan yang dikenal sebagai *transfer learning* menggunakan model yang telah dilatih terhadap dataset sebelumnya untuk menyelesaikan masalah serupa. Metode ini memulai dengan dataset sebelumnya, kemudian mengubah parameternya agar sama dengan dataset baru [10]. Sebuah *model pretrained* yang telah dilatih sebelumnya dibutuhkan dalam metode *transfer learning*. Dalam FER, model *pretrained* yang memenuhi persyaratan ini adalah *keras\_vggface*. Arsitektur untuk model ini adalah *resnet50*, *senet50* [5], dan *vgg16*. Ketiga arsitektur ini dapat membantu dalam proses pembuatan model pengenalan ekspresi wajah [11].

Pekerjaan yang diusulkan mempertimbangkan efisiensi sistem rekognisi citra dengan merancang sebuah arsitektur ringan untuk pengenalan ekspresi wajah dengan menerapkan teknik pengecilan kanal pada bagian tengah blok untuk mereduksi pemakaian parameter dan kompleksitas komputasi. Ini juga memanfaatkan deteksi wajah untuk pengenalan ekspresi pada kasus nyata, dan perhitungan evaluasi yang diharapkan akan membantu dalam sistem pemantauan keamanan gedung. Diharapkan penelitian ini akan meningkatkan efisiensi waktu dan kemudahan mengenali ekspresi wajah.

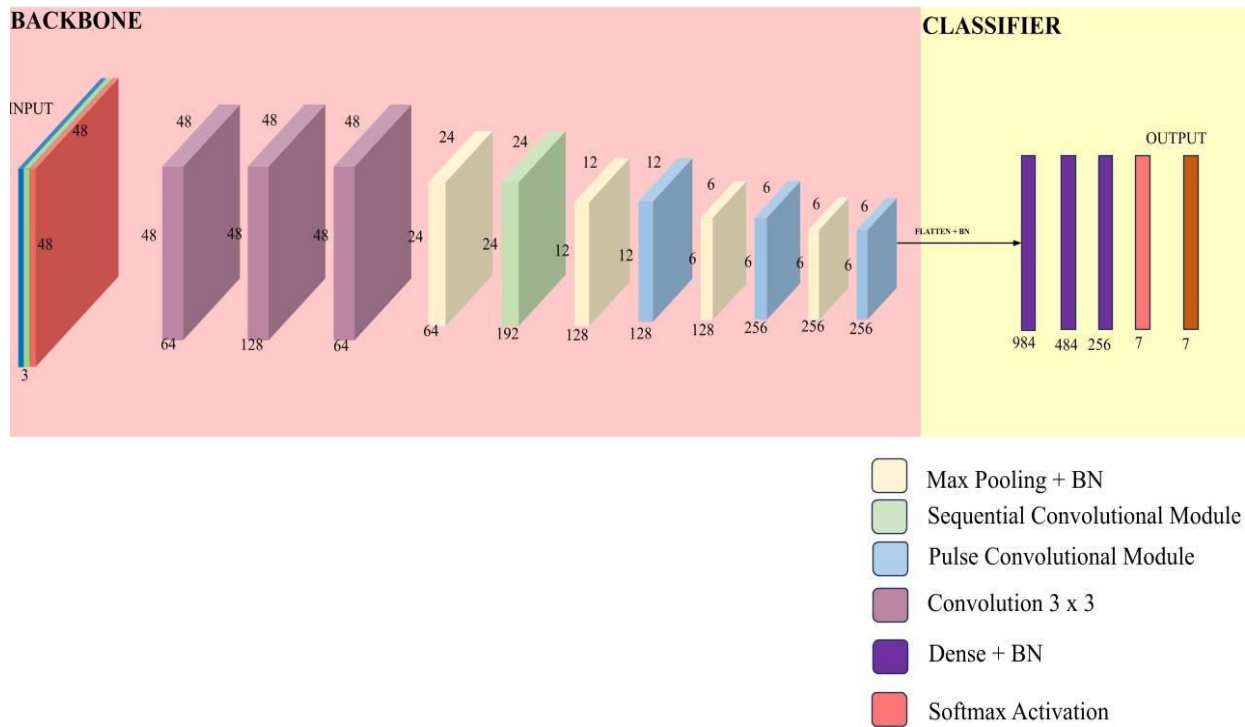
## II. METODE

### A. Backbone

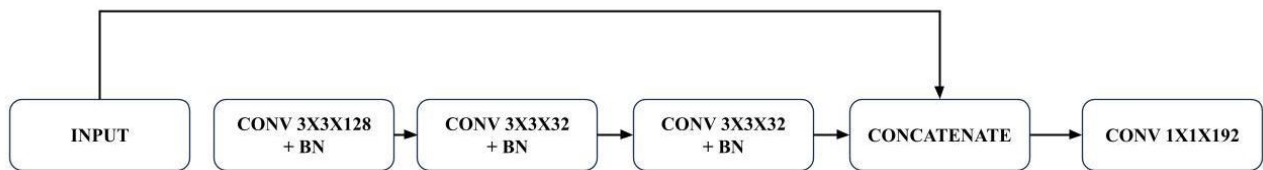
Tulang punggung atau backbone dalam arsitektur *convolutional neural networks* (CNN) sangat penting untuk mengekstraksi peta fitur dari ekspresi wajah [12]. Salah satu arsitektur terkenal dalam domain ini adalah VGG, yang diperkenalkan oleh tim Visual Geometry Group dari University of Oxford dalam makalah mereka yang berjudul "*Very Deep Convolutional Networks for Large-Scale Image Recognition*." Arsitektur ini dikenal dengan penggunaan kernel berukuran 3x3 di banyak lapisan, yang memungkinkan jaringan untuk menangkap fitur-fitur halus dalam gambar. Selain itu, VGG menggunakan *max pooling* untuk mereduksi dimensi peta fitur secara efisien, memungkinkan pengolahan yang lebih cepat tanpa kehilangan informasi penting. Nama VGG diambil dari jumlah lapisan yang dapat di-*update* bobotnya, menunjukkan kedalaman jaringan yang memungkinkan representasi yang lebih kompleks dan akurat dari data *input* [13]. Keseluruhan arsitektur dapat ditunjukkan pada Gambar 2. Modul sekuensial diusulkan untuk mengekstraksi fitur secara ringan. Ini menerapkan teknik pengecilan kanal pada bagian tengah untuk menekan penggunaan parameter berlebihan [14], seperti yang ditunjukkan pada Gambar 3.

### B. Modul Pengklasifikasi

Sebelum proses klasifikasi, peta fitur yang dihasilkan oleh jaringan harus diubah menjadi vektor menggunakan operasi *flatten*. Pada arsitektur VGG, klasifikator (*classifier*) terdiri dari dua lapisan *fully connected*, masing-masing dengan 4096 unit neuron, yang kemudian diakhiri dengan lapisan *softmax* untuk menghasilkan output klasifikasi. Struktur ini dirancang untuk mendukung jaringan dalam melakukan klasifikasi gambar setelah fitur-fitur penting diekstraksi oleh *backbone* [15]. Dengan demikian, kombinasi antara peta fitur yang di-*flatten* dan klasifikator yang kuat memastikan bahwa jaringan dapat secara efektif mengenali dan mengkategorikan berbagai jenis gambar berdasarkan fitur-fitur yang telah dipelajari. *Flatten* ditugaskan untuk membentuk peta fitur dari tulang punggung menjadi basis vektor. Kemudian *fully connected* memanfaatkan peran *neural network* mempekerjakan bobot-bobot terlatih dalam dimensi vektor untuk membentuk tujuh pengklasifikasi ekspresi wajah. Pada bagian akhir, fungsi *softmax* dipekerjakan untuk menghasilkan prediksi probabilitas dari masing-masing kelas.



Gambar 2. Arsitektur convolutional neural network yang diusulkan

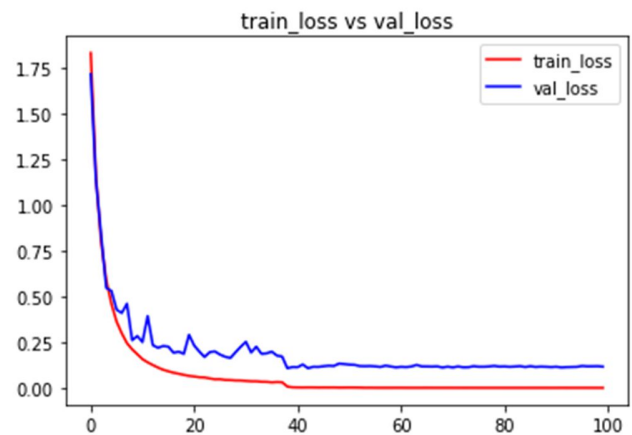


Gambar 3. Modul sekuensial konvolusi yang diusulkan

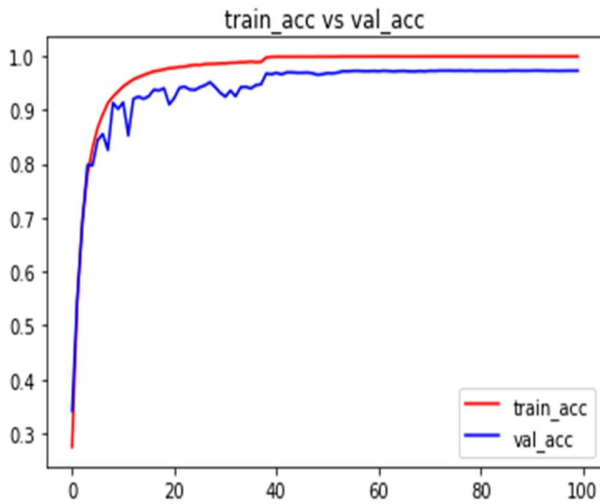
### C. Fungsi Kerugian

Pembelajaran mesin membutuhkan fungsi *loss* untuk mengukur perbedaan selisih prediksi dengan label kelas yang sebenarnya. Ini membantu model untuk mempelajari kesalahan-kesalahan dalam proses generasi bobot pada setiap neuron. Perhitungan kerugian dilakukan pada akhir backpropagasi yang mendukung fungsi *diferensial*. Selain itu, fungsi ini mendukung pengoptimalisasi proses *update* bobot yang memaksa jaringan untuk bekerja secara efektif.

Model yang diusulkan menerapkan fungsi *categorical cross entropy* [16] untuk mengevaluasi prediksinya. Fungsi ini mendorong jaringan untuk menemukan efektifitas pengukuran melalui pemberdayaan operasi logaritma *natural*. Fungsi ini menekan nilai yang kecil untuk setiap prediksi kelas yang benar dan memberikan penalti yang besar untuk prediksi yang keliru.



Gambar 4. Grafik hasil fungsi kerugian dari pelatihan model



Gambar 5. Grafik hasil akurasi dari pelatihan model

### III. DATASET, KONFIGURASI, PELATIHAN DAN PENGUJIAN

#### A. Dataset

Dataset dibutuhkan oleh jaringan yang berguna sebagai sumber pengetahuan fitur. Pengenalan ekspresi wajah manusia tergolong dalam sebuah sistem klasifikasi yang hanya membutuhkan label kelas ekspresi. Pekerjaan ini menggunakan KDEF dataset (*Karolinska Directed Emotional Faces*) yang terdiri atas tujuh kategori emosi manusia. Dataset ini berbasis pengambilan data laboratorium yang meminimalisasi gangguan latar belakang. Proses pelatihan dan pengujian melibatkan 4.900 citra berwarna yang mengagaskan pria dan wanita sebagai objek wajahnya. Untuk memperkaya variasi data, ini menerapkan teknik augmentasi secara berurutan. Teknik rotasi, flip, distorsi warna dan pencahayaan digunakan untuk menghasilkan citra sampel yang beragam. Proses ini menguntungkan jaringan untuk memperoleh informasi fitur yang lebih kaya dan dapat meningkatkan performa pelatihan.

#### B. Konfigurasi Pelatihan dan Pengujian

Untuk mengoptimalkan fase pelatihan dan menghindari masalah *overfitting*, jaringan yang diusulkan dilatih dalam konfigurasi tertentu. Untuk optimalisasi, diperlukan akselerator grafis seperti GTX 1080 TI. Model dievaluasi pada dataset *Karolinska Directed Emotional Faces* (KDEF). Selama pelatihan, digunakan *categorical cross-loss* dengan *epsilon*  $10^{-7}$  dan bobot acak pada awalnya. Optimasi dilakukan menggunakan *Adaptive Moment Estimation* (Adam) untuk *backpropagation*, dengan tingkat pembelajaran awal  $10^{-4}$  yang dapat disesuaikan jika akurasi tidak membaik setelah 10 *epoch*. Pelatihan dilakukan sebanyak 100 *epoch* dengan ukuran *batch* 16 untuk setiap dataset. Hasil inferensi menunjukkan kecepatan model diuji pada Jetson Nano 4GB. Adapun hasil proses pelatihan dari konfigurasi ini ditunjukkan melalui grafik akurasi dan fungsi kerugian pada Gambar 4 dan Gambar 5.

TABEL 1 HASIL PERBANDINGAN MODEL EKSPRESI WAJAH DENGAN ARSITEKTUR LAINNYA

No	Arsitektur	Parameter	100 Epoch		Test Real Application		
			Val. Acc	Val. Loss	FPS Total	FPS Class	FPS Detection
1	LeNet	528,311	0,7517	0,9728	4,95	5,23	96,51
2	MobileNetV1	3,246,983	0,9270	0,3570	4,08	4,28	84,65
3	MobileNetV2	1,906,275	0,9301	0,3196	4,14	4,33	95,91
4	MobileNetV2+BN	1,908,456	0,9205	0,3362	4,15	4,36	93,81
5	MobileNetV3Large	5,127,839	0,6928	18,4830	-	-	-
6	MobileNetV3Small	2,949,663	0,6812	21,7100	-	-	-
7	VGG 11 + BN	28,137,607	0,9622	0,2384	2,88	2,97	93,1
8	VGG 13 + BN	28,322,119	0,9698	0,1712	2,68	2,76	9249
9	VGG 16 + BN	33,631,815	0,9714	0,1882	2,54	2,61	95,04
10	VGG 19	38,941,511	0,9354	0,3641	2,32	2,38	94,45
11	Alexnet	44,023,175	0,1403	19,4600	2,54	2,61	93,55
12	Resnet 18	11,190,947	0,9051	0,4124	4,02	4,2	93,78
13	Resnet 34	16,586,531	0,8844	0,5067	3,37	3,5	92,08
14	GoogleNet	5,981,751	0,9223	0,8489	4,04	4,22	90,57
15	Resnet 50	21,310,243	0,8831	0,4866	3,37	3,5	90,69
16	Alexnet + BN	44,025,607	0,9179	0,9081	2,55	2,63	89,41
18	<b>Model yang diusulkan</b>	<b>2,893,395</b>	<b>0,9732</b>	<b>0,1330</b>	<b>15</b>	<b>18</b>	<b>78</b>

### IV. HASIL DAN PEMBAHASAN

#### A. Hasil dan Perbandingan Kinerja Dengan Metode Lain

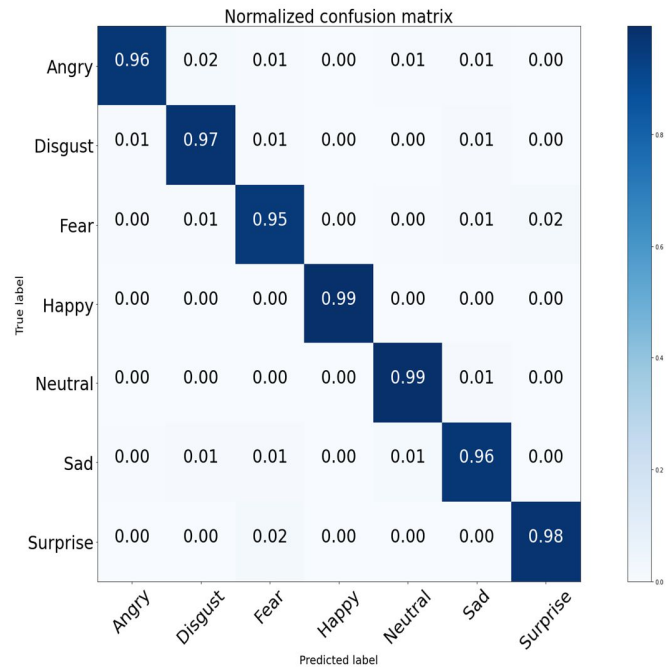
Bagian ini mengevaluasi hasil performa model yang diuji pada dataset citra berwarna KDEF. Ini mengukur seberapa akurat model yang membandingkan prediksi benar dengan jumlah seluruh sample uji. Berdasarkan Tabel 1, model yang kami usulkan menunjukkan performa yang lebih unggul dibandingkan dengan berbagai metode CNN lainnya. Model ini mencapai akurasi sebesar **97,32%**, yang 0,18% lebih tinggi dibandingkan dengan *VGG 16* yang menggunakan *Batch Normalization*. Selain itu, model kami memiliki jumlah parameter yang jauh lebih sedikit, yaitu hanya **2,8 juta**, dibandingkan dengan 33,6 juta parameter yang dimiliki oleh *VGG 16*. Efisiensi ini menunjukkan bahwa model kami tidak hanya lebih akurat tetapi juga lebih ringan dalam hal kompleksitas parameter. Matriks konfusi dari setiap prediksi kelas ditampilkan pada Gambar 6. Jika dibandingkan dengan metode CNN yang dirancang untuk perangkat *mobile* seperti

*MobileNet*, model kami tetap menunjukkan hasil yang optimal. Meskipun memiliki jumlah parameter yang lebih rendah, model ini mampu mencapai tingkat akurasi yang tinggi. Hal ini menjadikannya pilihan yang sangat tepat untuk implementasi pada perangkat dengan keterbatasan sumber daya, karena tidak mengorbankan kualitas performa meskipun lebih efisien dari segi ukuran model. Adapun hasil visualisasi prediksi pengenalan ekspresi wajah dari model yang diusulkan ditampilkan pada Gambar 7.

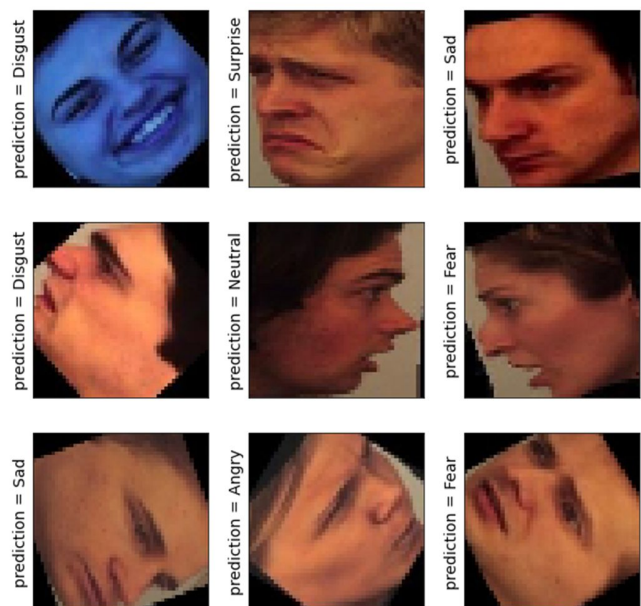
### B. Pengujian Efisiensi Model

Pendekatan visi komputer menuntut sebuah sistem penglihatan untuk dapat bekerja akurat, presisi dan memiliki waktu prosesi yang rendah [16]. Model CNN membuktikan sebagai sebuah ekstraksi fitur yang handal dan dapat secara efektif memisahkan fitur objek dan latar belakang. Namun, metode ini cenderung menghadirkan modul ekstraksi dengan parameter yang masih. Bahkan arsitektur-arsitektur populer merumuskan struktur dan penggunaan lapisan yang dalam yang melibatkan lapisan *neuron* yang berlimpah. Pekerjaan ini mengklaim penggunaan blok konvolusi yang panjang, dan jumlah lapisan kanal yang gemuk cenderung menghasilkan akurasi yang memuaskan pada dataset dengan jumlah kategori yang besar. Ini berdampak pada penggunaan bobot neuron yang banyak dan merepresentasikan waktu komputasi yang lebih besar juga. Sehingga cenderung memperlambat proses inferensi. Oleh karena itu, kebutuhan pengukuran kecepatan model pada tahap inferensi menjadi tahap signifikan untuk mengukur efisiensi dari sebuah perancangan sistem visi.

Tabel 1 menunjukkan perbandingan kecepatan model yang diuji pada perangkat berbiaya murah. Pekerjaan ini memanfaatkan perangkat Jetson Nano yang cenderung digunakan sebagai sebuah perangkat keras pada aplikasi robotika dan IoT. Perangkat ini memiliki spesifikasi GPU 4GB dengan sistem operasi Ubuntu yang membantu pengoptimalan sistem visi yang dirancang. Pengujian dilakukan melalui sebuah kamera webcam sebagai masukan *live streams* sehingga proses ini merepresentasikan sistem visi telah siap diimplementasikan pada kasus nyata. Selain itu, sistem pengenalan ekspresi wajah yang dirancang membutuhkan sebuah sistem deteksi wajah yang bertugas memisahkan daerah wajah dari latar belakang. Ini bertujuan untuk meningkatkan fokus kinerja dari sistem klasifikasi ekspresi wajah yang tidak akan terganggu oleh objek lainnya. Hasil dari deteksi wajah berupa patch wajah yang disesuaikan ukurannya menjadi 64x64 yang sesuai dengan ukuran masukan model. Selanjutnya, patch wajah ini diteruskan pada sebuah model yang telah dilatih untuk diprediksi kategori dari ekspresi wajahnya. Sebagai hasil, model yang diusulkan meraih kecepatan sebesar **15 FPS**. Ini mengungguli arsitektur-arsitektur populer lainnya, seperti *VGG*, *ResNet*, *AlexNet*, dan *GoogleNet*. Bahkan model yang diusulkan memperoleh kecepatan yang lebih baik dari arsitektur ringan lainnya, seperti *MobileNet*, dan *LeNet*.



Gambar 6. Hasil matriks konfusi dari prediksi model pada dataset KDEF



Gambar 7. Hasil prediksi citra ekspresi wajah pada dataset KDEF

Investigasi lainnya mengantarkan pada pengujian jumlah parameter dari model yang dihasilkan. Ini membandingkan banyaknya bobot neuron yang dipakai oleh jaringan, termasuk dengan penggunaan parameter pada normalisasi lapisan. Model yang diusulkan menghasilkan parameter yang lebih rendah dari model *VGG*, *ReNet* dan *GoogleNet*. Di Sisi yang lain, perbandingan dengan model *lightweight* yang populer menunjukkan bahwa model yang diusulkan memperoleh parameter yang lebih rendah dari *MobileNetV1* dan *MobileNetV3*.

## V. KESIMPULAN DAN SARAN

Kebutuhan deteksi ekspresi wajah yang ringan mengantarkan pada pekerjaan ini yang mendesain sebuah arsitektur CNN yang ringan untuk mengatasi masalah komputasi yang masif. Ini melibatkan operasi deretan konvolusi dengan lapisan yang dangkal untuk menekan penggunaan parameter yang berlebihan. Arsitektur yang diusulkan menerapkan *sekuensial operasi konvolusi* yang mengompresi kanal pada blok tengah. Hal ini menunjukkan bahwa arsitektur yang diusulkan lebih ringan dibandingkan arsitektur CNN yang populer lainnya. Strategi ini tidak mereduksi akurasi dari rekognisi ekspresi wajah yang menunjukkan hasil performa yang mengungguli dari para pesaing. Bahkan hasil ini mengilustrasikan model yang diusulkan lebih akurat dibandingkan arsitektur yang berat. Efisiensi model menguntungkan kontribusi riset yang menunjukkan parameter yang lebih rendah dan kecepatan yang lebih tinggi dari para pesaing. Pengembangan ke depan memungkinkan untuk menerapkan modul penguatan pada ekstraksi fitur untuk meningkatkan performa lebih tinggi.

## VI. KUTIPAN

- [1] B. C. L. Adiatma, E. Utami, dan A. D. Hartanto, “PENGENALAN EKSPRESI WAJAH MENGGUNAKAN DEEP CONVOLUTIONAL NEURAL NETWORK,” *EXPLORE*, vol. 11, no. 2, hlm. 75, Jul 2021, doi: 10.35200/explore.v11i2.478.
- [2] A. L. Sigit Gunto, E. Julianto, dan D. Budiyo, “Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network.”
- [3] S. Li dan W. Deng, “Deep Facial Expression Recognition: A Survey,” Apr 2018, doi: 10.1109/TAFFC.2020.2981446.
- [4] “[04] Adaptive Deep Metric Learning For Identity-Aware FER\_X Liu\_BVKV Kumar\_J You”.
- [5] W. J. Baddar dan Y. M. Ro, “Mode Variational LSTM Robust to Unseen Modes of Variation: Application to Facial Expression Recognition,” Nov 2018, [Daring]. Tersedia pada: <http://arxiv.org/abs/1811.06937>
- [6] W. Yang, J. Yu, T. Chen, Z. Liu, X. Wang, dan J. Shen, “Multi-threshold Deep Metric Learning for Facial Expression Recognition,” Jun 2024, [Daring]. Tersedia pada: <http://arxiv.org/abs/2406.16434>
- [7] L. M. Rasdi Rere, Sunarto Usna, dan Soegijanto, “Studi Pengenalan Ekspresi Wajah Berbasis Convolutional Neural Network,” *Seminar Nasional Teknologi Informasi dan Komunikasi STI&K*, vol. Volume 3, 2019.
- [8] “IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK DALAM DETEKSI EMOSI MANUSIA BERDASARKAN EKSPRESI WAJAH”.
- [9] “Waktu Reaksi dan Akurasi dalam Pengenalan Ekspresi Wajah”.
- [10] Samuel Sena, “Pengenalan Deep Learning Part 7: Convolutional Neural Network (CNN).”
- [11] K. Simonyan dan A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep 2014, [Daring]. Tersedia pada: <http://arxiv.org/abs/1409.1556>
- [12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, dan D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *Int J Comput Vis*, vol. 128, no. 2, hlm. 336–359, Feb 2020, doi: 10.1007/s11263-019-01228-7.
- [13] S. Woo, J. Park, J.-Y. Lee, dan I. S. Kweon, “CBAM: Convolutional Block Attention Module,” 2018, hlm. 3–19. doi: 10.1007/978-3-030-01234-2\_1.
- [14] T. Onisawa dan S. Kitazaki, “Recognition of Facial Expressions and Its Application to Human Computer Interaction,” dalam *INTELLIGENT BIOMETRIC TECHNIQUES in FINGERPRINT and FACE RECOGNITION*, Boca Raton: Routledge, 2022, hlm. 423–458. doi: 10.1201/9780203750520-13.
- [15] Z. Zhao, Q. Liu, dan S. Wang, “Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild,” *IEEE Transactions on Image Processing*, vol. 30, hlm. 6544–6556, 2021, doi: 10.1109/TIP.2021.3093397.
- [16] M. D. Putro, D.-L. Nguyen, dan K.-H. Jo, “A Fast CPU Real-Time Facial Expression Detector Using Sequential Attention Network for Human–Robot Interaction,” *IEEE Trans Industr Inform*, vol. 18, no. 11, hlm. 7665–7674, Nov 2022, doi: 10.1109/TII.2022.3145862.



**Reynold Frankie Robot**, anak pertama dari 4 bersaudara. Lahir di Balikpapan, 20 Februari 1979. Penulis menempuh pendidikan SMU Negeri 1 Balikpapan (1994-1997), melanjutkan pendidikan S1 Teknik Elektro di Universitas Kristen Satya Wacana (1997-2004), kemudian lanjut pendidikan S2 Magister Teknik Elektro Instrumentasi dan Kendali di Universitas Gajah Mada (2006-2008), dan sekarang menjadi staff pengajar di Fakultas Teknik Jurusan Teknik Elektro Program Studi Teknik Elektro, secara khusus berada dalam Kelompok Dosen Keahlian (KDK) Teknik Kendali dan Kecerdasan Buatan yang berfokus pada Otomasi, IoT, Komputer Visi, dan Robotika.