



Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network (CNN)

Richard Steven Immanuel Sihombing¹, Rafif Nauval Tuah Siregar², Vijay Sitorus³,
Timotius Selar Sitompul⁴

^{1,2,3,4}Universitas Negeri Medan

Alamat: Kampus: Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten
Deli Serdang, Sumatera Utara 20221

Korespondensi penulis: sihombingrichard106@email.com¹

Abstract. Facial expression recognition is an important research area in the advancement of machine learning. This research uses Convolutional Neural Network (CNN) as a method for recognizing facial expressions with a fairly high level of accuracy. This research uses a dataset obtained from Kaggle in the form of images of facial expressions, including surprised, happy, sad, afraid, angry and neutral. The MobilenetV2 CNN model was trained and tested using this dataset. The research results show that the model is able to recognize facial expressions with 78% accuracy on test data. It can be concluded that the MobilenetV2 model is quite capable of recognizing facial expressions.

Keywords: facial expressions, CNN, MobilenetV2

Abstrak. Pengenalan ekspresi wajah merupakan bidang penelitian yang penting dalam kemajuan pembelajaran mesin. Penelitian ini menggunakan Convolutional Neural Network (CNN) sebagai metode untuk mengenali ekspresi wajah dengan tingkat akurasi yang cukup tinggi. Penelitian ini menggunakan dataset yang diperoleh dari Kaggle berupa citra ekspresi wajah, termasuk terkejut, bahagia, sedih, takut, marah dan netral. Model CNN MobilenetV2 dilatih dan diuji menggunakan dataset ini. Hasil penelitian menunjukkan model mampu mengenali ekspresi wajah dengan akurasi 78% pada data uji. Dapat disimpulkan bahwa model MobilenetV2 cukup bisa mengenali ekspresi wajah.

Kata kunci: Ekspresi Wajah, CNN, MobilenetV2

LATAR BELAKANG

Wajah merupakan komponen tubuh yang dapat mencerminkan perasaan seseorang melalui ragam ekspresinya, apakah itu kegembiraan, kesedihan, kemarahan, ketakutan, kejutan, atau keadaan yang netral (Shafira, dkk., 2019). Dengan memahami kondisi emosional individu pada suatu waktu dan situasi tertentu, kita mampu mengevaluasi kesejahteraan psikologis mereka. Tambahan pula, ekspresi wajah memiliki dampak yang signifikan dalam mengidentifikasi perasaan seseorang dibandingkan dengan suara atau gerakan tubuh lainnya. (Samadiani, et al., 2019).

Ekspresi wajah berfungsi sebagai tanggapan alamiah yang mencerminkan perasaan atau emosi individu saat berinteraksi dengan suatu situasi tertentu. Dalam komunikasi antar manusia, ekspresi wajah menjadi elemen integral. Pengenalan ekspresi wajah menjadi keterampilan kunci untuk membangun hubungan interpersonal yang sehat,

dan menjadi fokus utama dalam studi perkembangan manusia, kesejahteraan psikologis, serta penyesuaian sosial. Tidak dapat disangkal bahwa pengenalan emosi memiliki peran sentral dalam mengalami empati, meramalkan perilaku prososial, dan dalam model kecerdasan emosional. Keahlian mengenali emosi dasar tampak sejak awal kehidupan, di mana anak-anak menggunakan ekspresi emosional sebagai petunjuk perilaku.

Kemajuan dalam teknologi deep learning saat ini, khususnya dalam konteks pelatihan, melibatkan penerapan transfer learning. Transfer learning merupakan suatu strategi dalam deep learning di mana model yang telah mengalami pelatihan untuk satu permasalahan dapat digunakan untuk menangani permasalahan lainnya. Pendekatan ini memungkinkan pencapaian tingkat akurasi yang tinggi dalam pelatihan deep learning, bahkan ketika terdapat keterbatasan pada jumlah sampel yang tersedia. Salah satu arsitektur awal yang sering digunakan dalam transfer learning adalah MobilenetV2.

MobileNet adalah salah satu struktur Convolutional Neural Network (CNN) yang dapat digunakan untuk mengatasi kebutuhan akan sumber daya komputasi yang terbatas (Apendi, dkk., 2023). Perbedaan mendasar antara arsitektur MobileNet dan CNN tradisional terletak pada penerapan layer konvolusi, di mana ketebalan filter disesuaikan dengan ketebalan input gambar (Sutama, dkk., 2020). MobilenetV2, versi kedua dari MobileNet, diperkenalkan pada bulan April 2017 dengan berbagai peningkatan (Mudzakir & Arifin, 2022). Versi ini memperkenalkan dua fitur tambahan, yaitu Linear Bottleneck dan Shortcut Connections antar Bottleneck (R. O. Ekoputris, 2018).

KAJIAN TEORITIS

Ekspresi Wajah

Manusia Mempunyai kemampuan untuk menampilkan emosi dalam bentuk ekspresi wajah. Terdapat enam ekspresi dasar wajah yang biasa digunakan oleh manusia yaitu : sedih, senang, marah, jijik, takut dan terkejut (Kobayashi, et al., 1995). Setiap ekspresi tersebut memiliki perbedaan pada tarikan otot-otot wajah(Tian, et al., 2001). Menurut Carole Wade & Carol Tavris (2007), salah satu gerak tubuh yang sering digunakan dalam proses komunikasi adalah ekspresi. Ekspresi wajah mencakup kombinasi beragam isyarat, di mana setiap isyarat memiliki makna dan dapat memengaruhi pesan verbal yang akan diungkapkan. Komunikasi juga dapat dipancarkan

melalui bahasa tubuh dan gerakan (sign language). Ada 6 Ekspresi wajah yang digunakan pada penelitian ini, yaitu:

- **Marah**
Tanda kemarahan umumnya tercermin dari kedua mata yang terlihat lebih tajam, alis yang mengkerut, menekan daerah di sekitar hidung, dan menyempitnya area di sekitar bibir.
- **Takut**
Ekspresi ketakutan dapat dikenali dari kedua alis yang naik secara bersamaan, kelopak mata bagian atas yang terangkat, ketegangan pada kelopak mata bagian bawah, dan bibir yang membentang secara horizontal menuju telinga.
- **Senang**
Ekspresi kebahagiaan ditandai dengan kerutan sudut mata, kedua sudut bibir yang menarik ke atas, dan kedua pipi yang terangkat. Ekspresi ini juga dikenal sebagai senyuman.
- **Netral**
Ekspresi netral adalah keadaan di mana tidak terdapat ekspresi emosi yang kentara, tidak menunjukkan perasaan negatif ataupun positif.
- **Sedih**
Ekspresi kesedihan menggambarkan mata bagian atas yang turun ke arah bawah, mata yang tampak tidak fokus, dan sudut bibir yang sedikit turun.
- **Kaget**
Ekspresi keterkejutan dicirikan oleh kedua alis yang naik, mulut yang sedikit terbuka, dan mata yang terbelalak.

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) tergolong dalam kategori deep neural network, mirip dengan neural network pada umumnya, yang terdiri dari sejumlah besar neuron memiliki bobot dan bias yang dapat diatur melalui proses pelatihan. Setiap neuron dalam CNN menerima berbagai input, menjalankan operasi komputasi, menghasilkan output, dan terhubung dengan neuron lainnya. Penggunaan istilah "konvolusi" mengacu pada operasi aljabar linear yang melibatkan perkalian matriks filter pada citra yang sedang diolah. Proses ini, yang dikenal sebagai lapisan konvolusi, merupakan salah satu jenis lapisan yang dapat ada dalam struktur jaringan CNN, meskipun sering dianggap

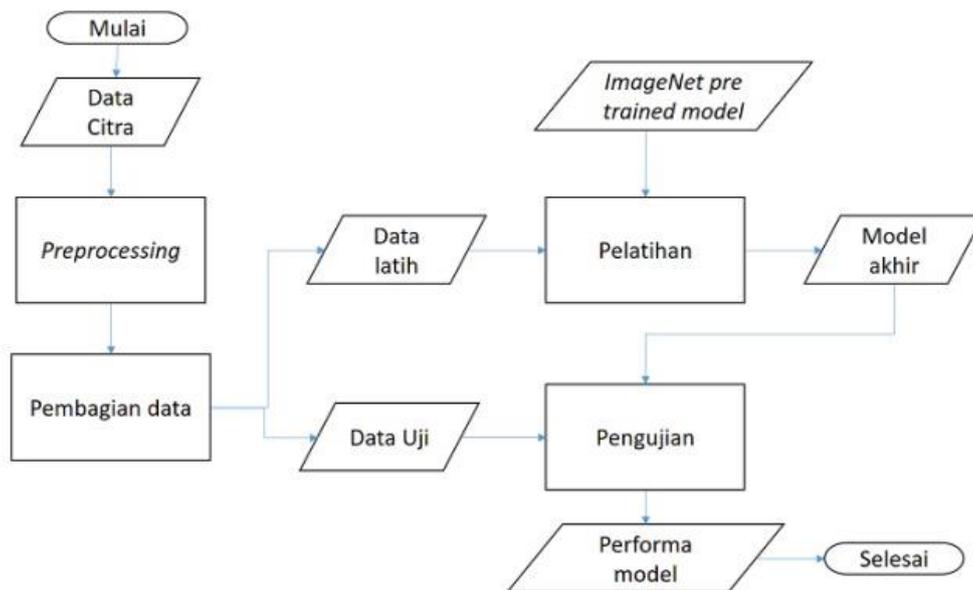
sebagai elemen paling vital. Lapisan lain yang umumnya digunakan melibatkan Pooling Layer, Normalization Layer, dan Fully Convolutional (FC) Layer. Keistimewaan CNN terletak pada asumsi eksplisit bahwa inputnya selalu berupa citra digital, mempermudah penyesuaian parameter arsitektur jaringan dan meningkatkan efisiensi secara keseluruhan. (Stanford University, 2016).

MobilenetV2

Setelah dilakukan pembagian data set, citra dilatih dengan menggunakan arsitektur mobilenetV2. MobilenetV2 adalah salah satu arsitektur jaringan saraf tiruan yang dirancang khusus untuk tugas pengenalan gambar pada perangkat bergerak atau mobile devices.

METODE PENELITIAN

Pada tahap ini berisi langkah-langkah penelitian yang penulis lakukan. Langkah-langkah ini berupa, pengumpulan dataset, lalu melakukan preprocessing citra, lalu melakukan ekstraksi fitur citra menggunakan imageNet pretrained model, lalu yang terakhir melakukan klasifikasi hasil ekstraksi fitur dengan menggunakan algoritma Convolutional Neural Network.



Gambar 1. Flowchart Alur Penelitian

Dataset

Dataset yang penulis gunakan adalah dataset yang dikumpulkan sendiri oleh peneliti menggunakan kamera handphone dan citra yang digunakan adalah citra dari para penulis sendiri. Dataset ini berisi kelas yang berkaitan dengan ekspresi wajah manusia yang berbeda, yaitu *Surprise, Anger, Happiness, Sad, Neutral, Fear*. Dataset ini terbagi menjadi dua yaitu data training sebanyak 123 sampel dan data testing sebanyak 24 sampel.

Preprocessing citra

Preprocessing adalah salah satu tahap penting yang dimana langkah-langkah yg dilakukan sebelum memasukkan citra kedalam model atau algoritma. Tujuannya adalah untuk mempersiapkan dan meningkatkan kualitas atau kecocokan citra dengan kebutuhan.

imageNet pretrained model

imageNet pretrained model merupakan model *Deep Learnig* yang telah dilatih dengan menggunakan dataset ImageNet. Model ini pada umumnya dilatih menggunakan Convolutional Neural Network (CNN) dan telah mencapai performa terbaik dalam pengklasifikasian citra. Model imageNet pretrained yang kami gunakan yaitu mobilenetV2.

Pengujian

Setelah hasil ekstraksi fitur didapat, maka langkah selanjutnya yaitu mengklasifikasikan menggunakan CNN. Pada penelitian ini model CNN yang akan digunakan adalah model mobilenetV2. Proses pengklasifikasian ekspresi wajah ini menggunakan CNN dengan varian model mobilenetV2 yang terdiri dari layer konvolusi standar dengan 32 filter kemudian diikuti dengan 19 layer *residual bottleneck*.

HASIL DAN PEMBAHASAN

Dataset yang berisi kumpulan citra ekspresi wajah dibagi kedalam 6 label class yaitu ekspresi Angry (marah), Fear (takut), Happy (senang), Neutral (netral), Sad (Sedih), Surprise (terkejut). Jumlah data pelatihannya sebanyak 22619 citra dan data validasinya sebanyak 5654 citra yang terbagi di 6 class yaitu angry (marah), fear (takut), happy (senang), neutral (netral), sad (sedih) surprise (terkejut).

1. Tahap Preprocessing

a. Batch Size

Batch Size merupakan parameter yang digunakan saat pelatihan model, Batch size berfungsi untuk menetapkan jumlah sampel penelitian yang diproses dalam satu iterasi yang mana pada deep learning dikenal dengan “epoch”.

b. Data Augmentasi

Data Augmentasi adalah Teknik untuk pengolahan citra yang digunakan untuk memperbanyak atau menggandakan dataset dengan menciptakan berbagai variasi dari data yang telah ada karena berfungsi untuk meningkatkan keragaman pada dataset sehingga model yang dibangun akan lebih memahami pola-pola dan menghasilkan akurasi yang lebih tinggi.

c. split data

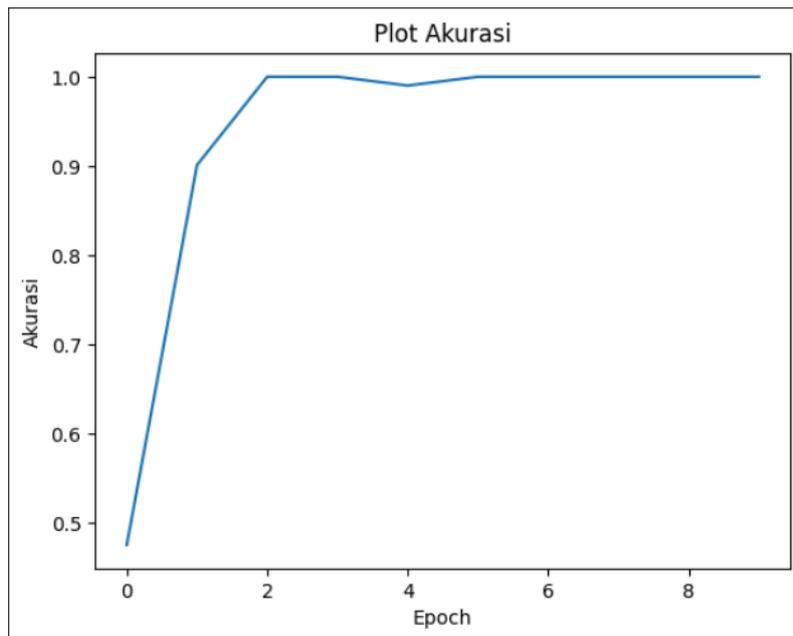
Split data adalah tahapan untuk membagi dataset kedalam 2 bagian yaitu data training dan data validation (testing). Data training digunakan agar model mempelajari, mengekstrak fitur, dan memahami pola dari dataset, sedangkan validation digunakan untuk model agar dapat melakukan prediksi atau identifikasi sebuah data kedalam labelnya. Pada penelitian ini perbandingan split data yang dilakukan yaitu 80% data training dan 20% data validation.

d. labelling

CNN adalah algoritma supervised learning, labelling ini bertujuan untuk memberikan nilai target kepada setiap citra dengan citra marah diberi label kategori 0, citra takut diberi label 1, citra senang diberi label 2, netral diberi label 3, citra sedih diberi label 4, dan citra terkejut diberi label 5. Proses ini dilakukan pada data training dan data testing, dan berfungsi sebagai dasar untuk pelatihan dan pengujian model CNN dalam mengenali gambar.

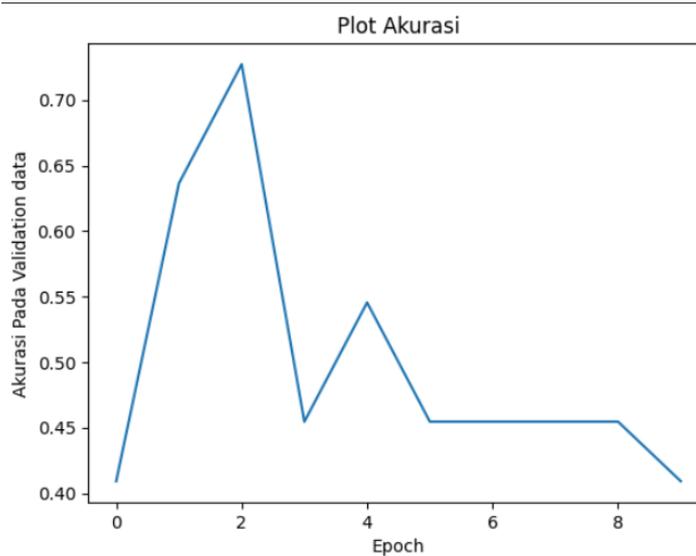
2. Hasil

Setelah selesai melalui proses pengujian sebanyak 10 epoch maka didapat hasil akurasi nya seperti gambar dibawah.



Gambar 2. Plot akurasi data Training

Gambar tersebut menunjukkan peningkatan akurasi yang tinggi dari epoch 1 ke 2 dan mengalami sedikit penurunan pada epoch 4 dan naik lagi sampai epoch 10 dan didapat hasil akurasi nya pada epoch 10 sebesar 100% dan untuk data validation didapatkan hasil yang kurang memuaskan yaitu sebesar 40%, walaupun saat awal pengujian sempat naik di 70%. Plot untuk akurasi data validasi ditampilkan plotnya seperti gambar dibawah ini.



Gambar 3. Plot Akurasi data validation

KESIMPULAN DAN SARAN

Penelitian ini mengimplementasi model mobilenetV2 dalam mengenali ekspresi wajah pada manusia yaitu marah, takut, senang, netral, sedih, dan terkejut. Hasil yang diperoleh dari menggunakan model tersebut yaitu akurasi sebesar 100%. Sedangkan pada data validation hasil yang diperoleh yaitu hanya sebesar 40%. Maka dari percobaan tersebut dapat dikatakan mobilenetV2 cukup baik digunakan untuk mengenali ekspresi wajah pada manusia.

DAFTAR REFERENSI

- S. S. Shafira, N. Ulfa, H. A. Wibawa and Rismiyati, "Facial Expression Recognition Using Extreme Learning Machine," 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2019, pp. 1-6, doi: 10.1109/ ICICoS48119.2019.8982443.
- S. Apendi, C. Setianingsih, and M. W. Paryasto, "Deteksi Bahasa Isyarat Sistem Isyarat Bahasa Indonesia Menggunakan Metode Single Shot Multibox Detector," eProceedings of Engineering, vol. 10, no. 1, 2023.
- V. A. Utama, S. A. Wibowo, and R. Rahmania, "Investigasi Pengaruh Step Training pada Metode Single Shot Multibox Detector untuk Marker dalam Teknologi Augmented Reality," Jurnal Ilmiah Fifo, vol. 12, no. 1, pp. 1–11, 2020.
- I. Mudzakir and T. Arifin, "Klasifikasi Penggunaan Masker dengan Convolutional Neural Network Menggunakan Arsitektur MobileNetv2," EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi, vol. 12, no. 1, pp. 76–81, 2022.
- R. O. Ekoputris, "MobileNet: Deteksi Objek pada Platform Mobile," Medium, May, vol. 9, 2018
- Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia. Algor, 2(1), 12-20.
- Guntoro, A. S., Julianto, E., & Budiyanto, D. (2022). Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network. Jurnal Informatika Atma Jogja, 3(2), 155-160.
- S. Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images," Int. J. Sci. Res. Publ., vol. 9, no. 10, p. p9420, 2019