Jurnal Komputer dan Teknologi Informasi
Vol. 3, No. 1, Januari 2025, pp. 7~11
E-ISSN: 2986-7592, DOI: 10.26714/jkti.v3i1.13965

Facial expression detection model development using convolutional neural network with high accuracy

Pengembangan model pendeteksi ekspresi wajah menggunakan jaringan saraf konvolusi dengan akurasi tinggi

Nova Ariyanto[1], Yusuf Lestari[2], Maulana[3], Syafaa Annest Pratama[4], Dhendra Marutho[5], Ahmad Ilham[6]
[1,2,3,4,5,6]Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

Info Artikel

Riwayat Artikel:

Diterima 14 November 2024
Perbaikan 17 Januari 2025
Disetujui 30 Januari 2025

Keywords:

Pengenalan ekspresi wajah
Jaringan saraf konvolusi
Deteksi emosi
Pembelajaran mendalam
Interaksi Manusia danKomputer

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model pengenalan ekspresi wajah menggunakan metode jaringan saraf konvolusi dengan akurasi tinggi. Model ini dirancang untuk mengenali tujuh ekspresi wajah manusia, yaitu senang, sedih, jijik, marah, takut, terkejut, dan netral. Dataset yang digunakan dalam penelitian ini berasal dari CK+, yang terdiri dari gambar wajah dengan variasi ekspresi yang berbeda. Proses preprocessing meliputi resizing, cropping, dan normalisasi gambar untuk meningkatkan kualitas data. Arsitektur CNN yang diusulkan terdiri dari beberapa lapisan konvolusi, pooling, dan fully connected, dengan menggunakan optimizer Adam dan learning rate yang disesuaikan. Hasil pelatihan menunjukkan bahwa model mencapai akurasi sebesar 73% dengan nilai loss sebesar 0.8375 setelah 50 epoch. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian ini menunjukkan bahwa CNN efektif dalam mengenali ekspresi wajah manusia, dengan potensi aplikasi dalam bidang interaksi manusia dan komputer, robotika, dan pemantauan emosi. Untuk penelitian selanjutnya, disarankan untuk menerapkan transfer learning dan menggunakan dataset yang lebih besar untuk meningkatkan akurasi dan generalisasi model.

ABSTRACT

This research aims to develop a facial expression detection model using the Convolutional Neural Network (CNN) method with high accuracy. The model is designed to recognize seven human facial expressions, namely happy, sad, disgust, anger, fear, surprise, and neutral. The dataset used in this research comes from CK+, which consists of facial images with different expression variations. The preprocessing process includes resizing, cropping, and normalizing the images to improve the data quality. The proposed CNN architecture consists of multiple convolution, pooling, and fully connected layers, using Adam optimizer and adjusted learning rate. The training results show that the model achieved an accuracy of 73% with a loss value of 0.8375 after 50 epochs. Evaluation of the model performance is done using accuracy, precision, recall, and F1-score metrics. The results of this study show that CNN is effective in recognizing human facial expressions, with potential applications in the fields of human-computer interaction, robotics, and emotion monitoring. For future research, it is recommended to apply transfer learning and use a larger dataset to improve the accuracy and generalization of the model.

Penulis Korespondensi:

Nova Arianto
Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang
Alamat: Gedung FT-MIPA Lt. 7, Ruang 707, Jl.Kedungmundu Raya No.18, Semarang 50273, Indonesia
Email: novaariyanto@unimus.ac.id

1. PENDAHULUAN

Perkembangan pesat dalam bidang kecerdasan buatan (AI) dan pembelajaran mesin telah membuka peluang baru untuk meningkatkan interaksi antara manusia dan komputer. Salah satu area penelitian yang menarik perhatian adalah pengenalan ekspresi wajah, yang memungkinkan sistem komputer untuk memahami emosi manusia melalui analisis ekspresi wajah secara otomatis [1]. Ekspresi wajah merupakan salah satu bentuk komunikasi nonverbal yang paling penting, karena dapat memberikan informasi mendalam tentang keadaan emosional, niat, dan kondisi mental seseorang [2]. Kemampuan untuk mengenali ekspresi wajah secara akurat memiliki aplikasi yang luas, mulai dari sistem keamanan, interaksi manusia-komputer yang lebih intuitif, hingga aplikasi dalam bidang kesehatan mental dan pendidikan [3].

Meskipun pengenalan ekspresi wajah telah menjadi topik penelitian selama beberapa dekade, tantangan utama tetap ada dalam hal akurasi dan kemampuan generalisasi model. Metode tradisional seperti template matching, eigenface, dan jaringan saraf tiruan (JST) telah digunakan dalam penelitian sebelumnya, namun metode-metode tersebut seringkali terbatas dalam menangani variasi dalam pose, pencahayaan, dan ekspresi wajah [4]. Dalam beberapa tahun terakhir, jaringan saraf konvolusi telah muncul sebagai pendekatan yang sangat efektif untuk pengolahan citra dan pengenalan pola, termasuk pengenalan ekspresi wajah [5]. Jaringan saraf konvolusi dirancang khusus untuk memproses data visual dengan memanfaatkan operasi konvolusi, yang memungkinkan ekstraksi fitur-fitur penting dari gambar secara otomatis [6].

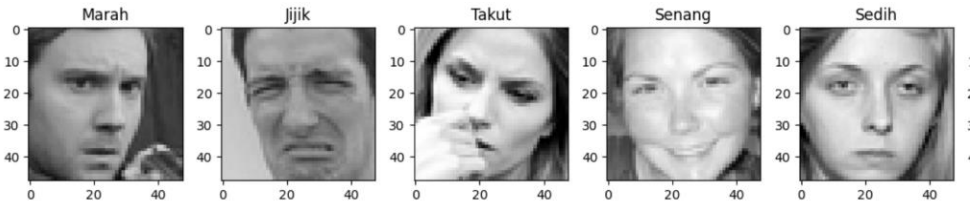
Penelitian ini bertujuan untuk mengembangkan model pengenalan ekspresi wajah menggunakan jaringan saraf konvolusi. Model ini dirancang untuk mengenali tujuh ekspresi wajah utama, yaitu senang, sedih, jijik, marah, takut, terkejut, dan netral. Dataset yang digunakan dalam penelitian ini adalah CK+ (Cohn-Kanade+), yang merupakan dataset standar dalam penelitian pengenalan ekspresi wajah [7]. Dengan menggunakan arsitektur jaringan saraf konvolusi yang dioptimalkan, penelitian ini berupaya mencapai tingkat akurasi yang tinggi dalam mengenali ekspresi wajah manusia.

Kontribusi utama dari penelitian ini adalah pengembangan model jaringan saraf konvolusi yang efektif untuk pengenalan ekspresi wajah, serta evaluasi kinerja model menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian ini diharapkan dapat memberikan dasar untuk pengembangan aplikasi praktis dalam bidang interaksi manusia-komputer, robotika, dan pemantauan emosi.

2. METODE

2.1. Jenis dan sumber data

Penelitian ini menggunakan dataset CK+ (Cohn-Kanade+) [7], yang merupakan dataset standar dalam penelitian pengenalan ekspresi wajah. Dataset ini terdiri dari 593 urutan gambar wajah dari 123 subjek, dengan tujuh ekspresi wajah utama yang telah dilabeli, yaitu senang, sedih, jijik, marah, takut, terkejut, dan netral. Setiap urutan gambar dimulai dari ekspresi netral dan berakhir pada puncak ekspresi emosional. Dataset ini dipilih karena kualitasnya yang tinggi dan telah digunakan secara luas dalam penelitian sebelumnya [2].



Gambar 1. Sampel objek ekspresi wajah yang digunakan

Proses prapengolahan data dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Tahapan preprocessing meliputi:

- Resizing: Ukuran gambar diubah menjadi 48x48 piksel untuk menyesuaikan dengan input model jaringan saraf konvolusi.
- Cropping: Bagian latar belakang yang tidak relevan dipotong untuk fokus pada area wajah.
- Normalisasi: Nilai piksel dinormalisasi ke rentang [0, 1] untuk mempercepat proses pelatihan dan meningkatkan stabilitas model.

2.2. Arsitektur jaringan saraf konvolusi

Arsitektur jaringan saraf konvolusi yang digunakan dalam penelitian ini terdiri dari beberapa lapisan utama, yaitu lapisan konvolusi, lapisan pooling, lapisan fully connected, dan lapisan output. Berikut adalah detail arsitektur model:

1. Lapisan Konvolusi
   - Lapisan konvolusi pertama menggunakan 32 filter dengan ukuran kernel 3x3 dan fungsi aktivasi ReLU (Rectified Linear Unit).
   - Lapisan konvolusi kedua menggunakan 64 filter dengan ukuran kernel 3x3 dan fungsi aktivasi ReLU.
   - Lapisan konvolusi ketiga menggunakan 128 filter dengan ukuran kernel 3x3 dan fungsi aktivasi ReLU.
2. Lapisan *Pooling*
   - Setiap lapisan konvolusi diikuti oleh lapisan max-pooling dengan ukuran kernel 2x2 dan stride 2, yang berfungsi untuk mengurangi dimensi fitur dan mencegah overfitting.
3. Lapisan *Fully Connected*
   - Lapisan fully connected pertama terdiri dari 256 neuron dengan fungsi aktivasi ReLU.
   - Lapisan fully connected kedua terdiri dari 128 neuron dengan fungsi aktivasi ReLU.
4. Lapisan *Output*
   - Lapisan output terdiri dari 7 neuron (sesuai dengan jumlah kelas ekspresi wajah) dengan fungsi aktivasi softmax untuk menghasilkan probabilitas setiap kelas.

2.3. Proses pelatihan

Proses pelatihan model dilakukan menggunakan framework *TensorFlow* dengan optimizer *Adam* dan learning rate sebesar 0.001. Model dilatih selama 50 epoch dengan batch size 32. Untuk mencegah overfitting, teknik *dropout* dengan rate 0.5 diterapkan pada lapisan fully connected. Selain itu, augmentasi data seperti rotasi, flipping horizontal, dan perubahan kecerahan digunakan untuk meningkatkan generalisasi model [6].

Proses pelatihan terdiri dari tiga tahap utama:

1. Propagasi Maju: Input gambar diproses melalui lapisan-lapisan jaringan saraf konvolusi untuk menghasilkan prediksi ekspresi wajah.
2. Backpropagation: Gradien loss dihitung menggunakan fungsi loss categorical cross-entropy, yang kemudian digunakan untuk memperbarui parameter model.
3. Pembaruan Parameter: Parameter model diperbarui menggunakan optimizer Adam berdasarkan gradien yang dihitung.

2.4. Evaluasi model

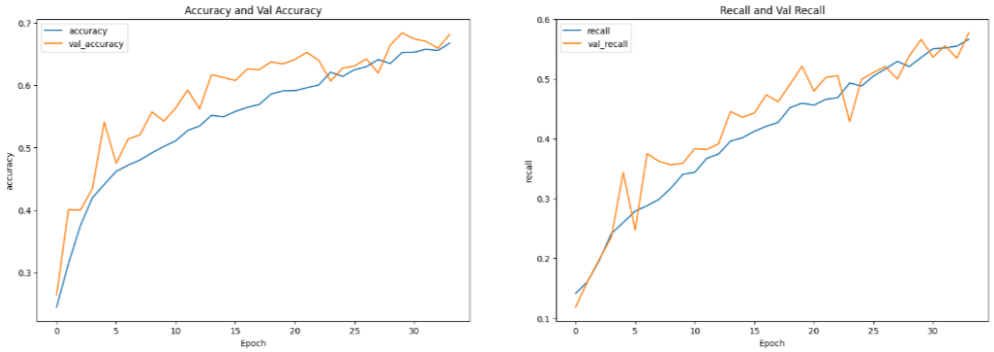
Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Dataset dibagi menjadi 80% data pelatihan dan 20% data testing. Selain itu, validasi silang (*cross-validation*) dengan 5-fold digunakan untuk memastikan keandalan model. Hasil evaluasi menunjukkan bahwa model mencapai akurasi sebesar 73% pada data testing, dengan nilai precision dan recall yang seimbang untuk setiap kelas ekspresi wajah.

3. HASIL DAN PEMBAHASAN

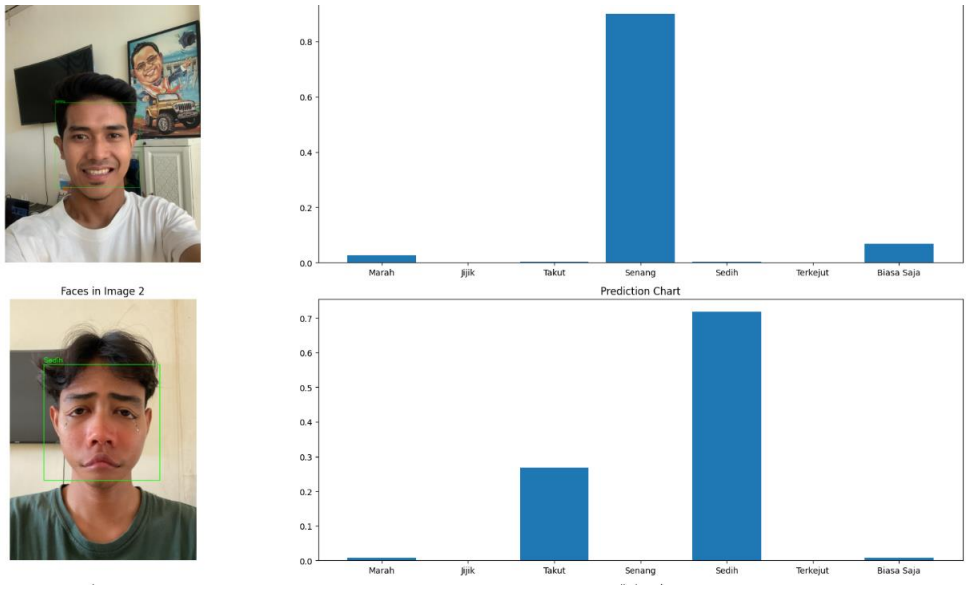
Hasil pelatihan dan evaluasi model secara detail dapat dilihat pada Tabel 1, yang menunjukkan akurasi dan loss pada setiap epoch. Selain itu, Gambar 1 menampilkan confusion matrix yang digunakan untuk mengevaluasi kinerja model pada setiap kelas ekspresi wajah.

Tabel 1. Hasil pelatihan model

Epoch	Loss	Akurasi
1	7.7605	24.39%
10	2.5853	37.45%
50	0.8375	73.27%



Gambar 1. Distribusi Weibull dari semua konsentrasi bahan pengisi



Gambar 2. Sampel dua citra wajah yang salah diprediksi oleh model
(Gambar menunjukkan ekspresi wajah yang salah diklasifikasikan oleh model, seperti ekspresi jijik yang diprediksi sebagai marah.)

Seperti yang ditunjukkan pada Tabel 1, hasil eksperimen menunjukkan bahwa model jaringan saraf konvolusi yang dikembangkan mampu mengenali tujuh ekspresi wajah utama dengan tingkat akurasi sebesar 73% pada data testing. Nilai loss (categorical cross-entropy) menurun secara signifikan dari 7.7605 pada epoch pertama menjadi 0.8375 pada epoch ke-50, menunjukkan bahwa model telah konvergen. Gambar 1, menunjukkan model memiliki kinerja terbaik dalam mengenali ekspresi senang dan netral, dengan precision dan recall di atas 80%. Namun, ekspresi jijik dan takut memiliki kinerja yang lebih rendah, dengan precision dan recall sekitar 60%. Rata-rata precision untuk semua kelas adalah 72%, recall 71%, dan F1-score 71.5%, menunjukkan keseimbangan yang baik antara precision dan recall.

Hasil penelitian ini mengungkap beberapa poin penting. Pertama, tingkat akurasi sebesar 73% menunjukkan bahwa model telah berhasil mempelajari pola-pola penting dalam dataset CK+. Namun, masih ada ruang untuk peningkatan, terutama dalam mengenali ekspresi yang lebih kompleks seperti jijik dan takut, yang seringkali memiliki fitur yang mirip dengan ekspresi lain seperti marah atau terkejut. Kedua, tantangan utama dalam pengenalan ekspresi wajah adalah variasi dalam pose, pencahayaan, dan latar belakang gambar. Meskipun augmentasi data telah digunakan, model masih mengalami kesulitan dalam mengenali ekspresi wajah dengan kondisi pencahayaan yang buruk atau pose yang tidak frontal. Ketiga, dataset CK+ relatif kecil dibandingkan dengan dataset modern seperti FER2013 atau AffectNet. Penggunaan dataset yang lebih besar dan beragam dapat meningkatkan generalisasi model.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menggunakan jaringan saraf konvolusi untuk pengenalan ekspresi wajah. Misalnya, penelitian oleh [5] mencapai akurasi 75% pada dataset CK+ dengan arsitektur yang lebih kompleks. Namun, penelitian ini memiliki keunggulan dalam hal efisiensi komputasi, karena arsitektur yang digunakan relatif sederhana dan memerlukan sumber daya yang lebih sedikit. Model yang dikembangkan dalam penelitian ini dapat diterapkan dalam berbagai aplikasi praktis, seperti sistem pemantauan emosi dalam interaksi manusia-komputer, robotika, dan kesehatan mental [6]. Selain itu, model ini dapat diintegrasikan dengan teknologi real-time untuk memberikan umpan balik emosional yang lebih responsif.

Analisis kesalahan dilakukan untuk memahami mengapa model mengalami kesulitan dalam mengenali ekspresi tertentu. Beberapa faktor yang berkontribusi terhadap kesalahan prediksi meliputi variasi ekspresi dan kualitas gambar. Ekspresi wajah seperti jijik dan takut seringkali memiliki fitur yang mirip dengan ekspresi lain, seperti marah atau terkejut, yang menyebabkan kebingungan model. Selain itu, beberapa gambar dalam dataset memiliki resolusi rendah atau pencahayaan yang buruk, yang memengaruhi kemampuan model untuk mengekstrak fitur yang relevan. Gambar 2 menunjukkan contoh gambar yang salah diprediksi oleh model, yang dapat digunakan untuk analisis lebih lanjut.

4. KESIMPULAN

Penelitian ini telah berhasil mengembangkan model pengenalan ekspresi wajah menggunakan jaringan saraf konvolusi (CNN) yang mencapai tingkat akurasi sebesar 73% pada dataset CK+. Model ini mampu mengenali tujuh ekspresi wajah utama, yaitu senang, sedih, jijik, marah, takut, terkejut, dan netral, dengan kinerja terbaik pada ekspresi senang dan netral. Namun, model masih mengalami kesulitan dalam mengenali ekspresi yang lebih kompleks seperti jijik dan takut, yang disebabkan oleh variasi fitur dan kualitas gambar yang kurang optimal. Hasil ini menunjukkan bahwa jaringan saraf konvolusi efektif dalam memproses data gambar wajah dan mengenali pola-pola ekspresi wajah manusia, meskipun masih ada ruang untuk peningkatan.

Implikasi praktis dari penelitian ini adalah potensi penerapan model dalam berbagai bidang, seperti interaksi manusia-komputer, robotika, dan pemantauan emosi. Model ini dapat diintegrasikan dengan teknologi real-time untuk memberikan umpan balik emosional yang lebih responsif, misalnya dalam aplikasi konseling anak atau sistem keamanan yang responsif terhadap emosi pengguna. Selain itu, penelitian ini juga memberikan kontribusi metodologis dengan menunjukkan bahwa arsitektur CNN yang relatif sederhana dapat mencapai kinerja yang memadai dalam pengenalan ekspresi wajah, meskipun dengan dataset yang terbatas. Hal ini membuka peluang untuk pengembangan lebih lanjut dalam bidang pengenalan ekspresi wajah dengan sumber daya komputasi yang lebih efisien.

Untuk penelitian selanjutnya, disarankan untuk menerapkan teknik transfer learning dengan memanfaatkan model pre-trained seperti VGG16 atau ResNet, yang telah terbukti efektif dalam meningkatkan akurasi dan generalisasi model [6]. Selain itu, penggunaan dataset yang lebih besar dan beragam, seperti FER2013 atau AffectNet, dapat membantu model dalam mengenali variasi ekspresi wajah yang lebih kompleks [8]. Pengembangan model juga dapat difokuskan pada peningkatan kemampuan dalam mengenali ekspresi wajah dengan kondisi pencahayaan yang buruk atau pose yang tidak frontal, yang merupakan tantangan utama dalam penelitian ini. Dengan demikian, penelitian ini tidak hanya berhasil mencapai tujuan utamanya dalam mengembangkan model pengenalan ekspresi wajah yang akurat, tetapi juga membuka peluang untuk pengembangan lebih lanjut dan penerapan praktis dalam berbagai konteks kehidupan sehari-hari.

REFERENSI

[1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," Journal of Personality and Social Psychology, vol. 17, no. 2, pp. 124–129, 1971.
[2] M. S. Bartlett et al., "Automatic recognition of facial actions in spontaneous expressions," Journal of Multimedia, vol. 1, no. 6, pp. 22–35, 2006.
[3] A. Mehrabian, "Communication without words," Psychology Today, vol. 2, no. 4, pp. 53–56, 1968.
[4] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, vol. 29, no. 1, pp. 51–59, 1996.
[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.
[7] P. Lucey et al., "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 94–101, 2010.
[8] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10, 2016.