

Pengenalan Ekspresi Wajah pada Gambar Beresolusi Rendah Menggunakan Metode Convolutional Neural Network Dengan Arsitektur EfficientNet

Im Abdul Karim¹, Mahmud Dwi Sulistiyono, S.T., M.T., Ph. D.², Dr. Ema Rachmawati, S.T., M.T.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹iimakariim@students.telkomuniversity.ac.id,

²mahmuddwis@telkomuniversity.ac.id,

³emarachmawati@telkomuniversity.ac.id

Abstrak

Ekspresi wajah merupakan saluran utama komunikasi non-verbal untuk menyampaikan emosi dan memiliki peranan penting dalam komunikasi manusia sehari-hari. Pengenalan ekspresi wajah berbasis citra menjadi topik penelitian yang cukup populer di bidang kecerdasan buatan dan visi komputer.

Namun, pengenalan ekspresi wajah pada citra input beresolusi rendah masih menjadi tantangan yang signifikan dan penelitian terkait hal tersebut masih relatif terbatas dalam perkembangan teknologi ini. Gambar beresolusi rendah seringkali ditemui dalam situasi pengambilan gambar dari sumber daya terbatas seperti CCTV atau perangkat sensor pada perangkat mobile, sehingga dapat mengakibatkan kehilangan detail penting pada wajah dan menyulitkan pengenalan ekspresi wajah dengan akurasi yang optimal. Pada penelitian ini membangun model pengenalan ekspresi wajah berbasis Convolutional Neural Network optimal yang mampu mengatasi kondisi citra input dengan resolusi rendah. Pemilihan arsitektur model EfficientNetB0 dengan melakukan teknik fine-tuning melatih seluruh lapisan dan teknik oversampling pada dataset CK+ untuk mengatasi kelas imbalance pada data latih. Hasil yang didapatkan dari penelitian ini dengan menganalisis pengaruh dari beberapa optimizer terhadap kinerja pada model yang diajukan diantaranya Adam, SGD, dan RMSProp dapat diambil hasil yang paling optimal yaitu optimizer RMSProp, dan di dapat nilai akurasi sebesar 96%, dengan *recall* sebesar 96%, *precision* 96,5%, dan *F1-Score* 95,84%. Berdasarkan hasil analisis nilai akurasi, *precision*, *recall*, dan *F1-Score* tinggi dimana model yang dibangun mampu mengoptimalkan kinerja pada sistem pengenalan ekspresi wajah terutama pada gambar beresolusi rendah.

Kata kunci : pengenalan ekspresi wajah, pembelajaran mendalam, jaringan syaraf konvolusi, gambar beresolusi rendah, efficientnet

Abstract

Facial expressions are the main channel of non-verbal communication to convey emotions and have an important role in everyday human communication. Image-based facial expression recognition has become a quite popular research topic in the fields of artificial intelligence and computer vision. However, recognizing facial expressions in low-resolution input images is still a significant challenge, and research regarding this is still relatively limited in the development of this technology. Low-resolution images are often encountered in shooting situations from limited resources such as CCTV or sensor devices on mobile devices, so they can result in the loss of important facial details and make it difficult to recognize facial expressions with optimal accuracy. In this research, we build an optimal Convolutional Neural Network-based facial expression recognition model that is able to overcome low-resolution input image conditions. Selection of the EfficientNetB0 model architecture by carrying out fine-tuning techniques to train all layers and oversampling techniques on the CK+ dataset to overcome the imbalance class in the training data. The results obtained from this research by analyzing the influence of several optimizers on the performance of the proposed model including Adam, SGD, and RMSProp, the most optimal results can be obtained, namely the RMSProp optimizer, and an accuracy value of 96%, with a recall of 96%, is obtained. precision 96,5%, and F1-Score 95,84%. Based on the results of the analysis of high accuracy, precision, recall and F1-Score values, the model built is able to optimize performance in facial expression recognition systems, especially in low resolution images.

Keywords: facial expression recognition, deep learning, convolutional neural network, low-resolution images, efficientnet

1. Pendahuluan

Latar Belakang

Di era perkembangan teknologi digital ini, pengenalan ekspresi wajah adalah salah satu bidang penelitian yang terus berkembang pesat di dunia kecerdasan buatan [1]. Ekspresi wajah adalah saluran utama komunikasi *non-verbal* untuk menyampaikan emosi dan memiliki peranan penting dalam komunikasi manusia sehari-hari.

Kebutuhan untuk mengenali ekspresi wajah semakin meningkat, dan telah terjadi peningkatan minat terhadap pengenalan ekspresi wajah yang memiliki penerapan aplikasi luas, termasuk bidang pemantauan keamanan [4][5], pendidikan [6], rehabilitasi medis [7], pengenalan ekspresi wajah di alam liar [8][9], dan berkendara yang aman [10].

Namun, untuk mengenali ekspresi wajah pada gambar beresolusi rendah masih menjadi sebuah tantangan di bidang pengenalan ekspresi wajah [2]. Gambar beresolusi rendah dapat ditemui dalam berbagai situasi termasuk pengambilan dari sumber daya terbatas, seperti kamera keamanan CCTV atau perangkat sensor yang terintegrasi dalam perangkat mobile. Gambar Beresolusi rendah sering kali mengandung informasi yang terbatas dan dapat mengakibatkan kehilangan detail penting pada wajah, sehingga mengurangi keakuratan dalam mengenali ekspresi wajah. Penelitian yang dilakukan Radhiyatul *et al*, gambar beresolusi rendah memiliki resolusi 64x64 piksel kebawah dan cenderung berdampak signifikan terhadap akurasi sistem[15]. Meskipun telah ada banyak penelitian yang mengeksplorasi pengenalan ekspresi wajah secara umum, fokus penelitian secara khusus pada gambar beresolusi rendah masih relatif terbatas.

Salah satu metode yang telah terbukti efektif dalam pengenalan citra, termasuk pengenalan ekspresi wajah adalah *Convolutional Neural Network* (CNN). CNN telah sukses diterapkan dalam berbagai tugas pengolahan citra digital, termasuk objek identifikasi dan ekstraksi fitur. Selain itu, CNN memiliki kinerja klasifikasi yang sangat baik dalam ekstraksi fitur gambar ekspresi wajah [3]. Oleh karena itu, alasan penggunaan metode *Convolutional Neural Network* (CNN) yaitu karena metode CNN merupakan salah satu algoritma *deep learning* yang mampu melakukan klasifikasi data dari jumlah data yang banyak dan dapat memperoleh nilai akurasi yang tinggi dalam konteks pengenalan ekspresi wajah. Terkait hal tersebut dalam konteks pengenalan ekspresi wajah, *optimizer* memegang peranan penting dalam mempengaruhi konvergensi dan akurasi model. *Optimizer* adalah algoritma yang digunakan untuk menyesuaikan bobot dalam jaringan neural selama proses pelatihan, yang pada akhirnya berdampak pada seberapa baik model tersebut mampu belajar dari data[14].

Penelitian yang dilakukan oleh Bodavarapu *et al*. dimana penelitian tersebut merancang sebuah sistem pengenalan ekspresi wajah untuk gambar beresolusi rendah menggunakan metode CNN dan *denoising techniques*. Hasil akurasi yang didapat pada penelitian tersebut menggunakan model arsitektur yang diusulkan yaitu FerConvNet dibandingkan dengan arsitektur *EfficientNetB7* terhadap dataset yang telah dibuat dengan menskalakan menjadi 48x48 piksel dimana pada penelitian tersebut mendefinisikan ukuran citra sebesar 48x48 piksel masuk kedalam kategori citra beresolusi rendah mencapai sebesar 71% pada arsitektur *FerConvNet* dan 65%. Dari hasil akurasi yang diperoleh model yang diusulkan tersebut tidak bekerja dengan baik dan menghasilkan akurasi yang tidak optimal [11]. Sementara penelitian yang dilakukan oleh Jose *et al*, dimana penelitian tersebut mengimplementasikan metode *voting residual network* atau arsitektur *resnet* dimana algoritma tersebut merupakan algoritma *deep learning* CNN untuk pengenalan ekspresi wajah terhadap gambar beresolusi rendah, dataset yang digunakan pada penelitian tersebut adalah CK+ yang mana dataset tersebut memiliki resolusi kepadatan sebesar 48x48 piksel dan dataset Affecnet yang diskalakan ulang menjadi 48x48 piksel untuk membandingkan hasil akurasi. Hasil yang diperoleh terhadap dataset CK+ dengan model yang diajukan yaitu sebesar 98% dan untuk dataset Affecnet mencapai 85%, dari hasil akurasi yang diperoleh model terhadap dataset CK+ membuktikan bahwa model yang diajukan bekerja dengan optimal dan dapat ditingkatkan lagi hasil kinerja dari modelnya [16]. Penggunaan teknik *transfer learning* dengan menggunakan arsitektur *EfficientNet* dan algoritma *boosting XGBoost* untuk membuat sistem pengenalan ekspresi wajah memperoleh hasil akurasi 99% terhadap dataset CK+, dari hasil yang diperoleh membuktikan bahwa penggunaan arsitektur *EfficientNet* dengan algoritma *XGBoost* dapat membantu untuk sistem pengenalan ekspresi wajah dimana model tersebut memiliki kinerja yang efisien dan dapat berkerja secara optimal [17].

Dalam penelitian ini, membangun sistem pengenalan ekspresi wajah pada gambar beresolusi rendah menggunakan metode *Convolutional Neural Network* (CNN) dengan memanfaatkan model arsitektur *EfficientNet* yang bertujuan untuk menganalisis pengaruh *optimizer* pada tingkat akurasi model yang diajukan terhadap gambar beresolusi rendah. *Optimizer* seperti SGD (Stochastic Gradient Descent), Adam, dan RMSprop memiliki karakteristik yang berbeda dalam proses pelatihan model, dan perbedaan tersebut dapat berdampak signifikan pada hasil akurasi akhir, terutama pada kondisi gambar beresolusi rendah. Dengan demikian, penelitian ini bertujuan untuk memahami bagaimana pilihan *optimizer* memengaruhi performa model CNN dalam konteks pengenalan ekspresi wajah pada gambar beresolusi rendah. Diharapkan penelitian ini dapat memberikan kontribusi positif terhadap pengembangan sistem pengenalan ekspresi wajah yang lebih handal dan efisien dibandingkan dengan penelitian sebelumnya, terutama dalam situasi di mana gambar beresolusi rendah.

Topik dan Batasannya

Berdasarkan latar belakang di atas, topik penelitian ini berangkat dari pertanyaan utama tentang bagaimana membangun sistem pengenalan ekspresi wajah dengan menggunakan arsitektur CNN *EfficientNet*, selain itu juga menganalisis pengaruh *optimizer* pada akurasi dalam gambar beresolusi rendah.

Berikut terdapat beberapa batasan masalah untuk penelitian ini, yaitu:

1. Dataset yang digunakan yaitu CK+ yang sudah diskala ulang menjadi 48 x 48 piksel, yang mana dataset tersebut masuk kedalam kategori *low-resolution images*[15].
2. Penelitian ini hanya berfokus pada klasifikasi 7 kelas dasar ekspresi wajah manusia seperti marah,

jijik, ketakutan, senang, sedih, terkejut, dan netral.

Tujuan

Tujuan dari penelitian Tugas Akhir ini adalah membangun sistem pengenalan ekspresi wajah dengan menggunakan arsitektur CNN *EfficientNet*, menganalisis pengaruh *optimizer* pada akurasi terutama dalam gambar beresolusi rendah.