

PENGENALAN EKSPRESI WAJAH SECARA REALTIME MENGGUNAKAN TRANSFER LEARNING PADA FACENET

Realtime Facial Expression Recognition Using Transfer Learning On FaceNet

Luky Vianika Sari*¹, Aziz Musthafa², Triana Harmini³

^{1,2,3} Universitas Darussalam Gontor

^{1, 2, 3} Ponorogo, Indonesia

e-mail: *lukyvianika@mhs.unida.gontor.ac.id, aziz@unida.gontor.ac.id,
triana@unida.gontor.ac.id

Abstrak - Ekspresi tubuh manusia merupakan bentuk respon alami yang menggambarkan suatu keadaan dan perasaan terhadap situasi tertentu. Pengenalan ekspresi wajah dapat dimanfaatkan dalam berbagai bidang. Seperti mengenali statistik perilaku konsumen bagi industri retail, dan menilai kepuasan individu pada suatu layanan. Namun, dalam mendeteksi ekspresi wajah terdapat sejumlah tantangan yang perlu diperhatikan salah satunya adalah diperlukan data yang cukup banyak agar komputer dapat mengenali gambar tersebut. Untuk mendapatkan data gambar yang banyak dibutuhkan waktu yang lama dalam mengumpulkan gambar menyesuaikan dengan kebutuhan. Tujuan dari penelitian ini adalah komputer mampu mendeteksi ekspresi wajah secara realtime tanpa perlu melibatkan data yang banyak. Oleh karena itu peneliti menggunakan metode transfer learning yang merupakan pemanfaatan model yang sudah dilatih oleh orang lain, yang selanjutnya digunakan dalam mendeteksi ekspresi wajah. Adapun arsitektur yang digunakan adalah FaceNet. FaceNet merupakan salah satu arsitektur deep learning yang digunakan untuk pengenalan, identifikasi, dan klasifikasi wajah yang selanjutnya akan dimodifikasi dengan menambahkan layer baru sehingga dapat mendeteksi ekspresi wajah. Hasil dari penelitian ini adalah akurasi data uji sebesar 93% dengan dataset sebanyak 150 gambar, terdiri dari 50 gambar marah, 50 gambar senang, dan 50 gambar netral.

Kata kunci - Ekspresi Wajah, FaceNet, Transfer Learning

Abstract - The expression of the human body is a form of natural response that describes a state and feelings to a particular situation. Facial expression recognition can be utilized in various fields. Such as recognizing consumer behavior statistics for the retail industry, and assessing the satisfaction of individuals on a service. However, in detecting facial expressions there are a number of challenges that need to be considered one of which is the need for enough data for the computer to recognize that image. To get a lot of image data it takes a long time in collecting images, according to need. The goal of the study was that computers were able to detect facial expressions in real time without the need to involve a lot of data. Therefore, researchers use transfer learning methods that are the utilization of models that have been trained by others, which are further used in detecting facial expressions. The architecture use in FaceNet. FaceNet is one of the deep learning architectures used for facial recognition, identification, and classification, it will then be modified by adding a new layer so that it can detect facial expressions. The result of this study was a 94% accuracy of test data with a dataset of 150 images, consisting of 50 angry images, 50 happy images, and 50 neutral images.

Keywords - Facial Expression, FaceNet, Transfer Learning

I. PENDAHULUAN

Pada bidang *Computer Vision* klasifikasi ekspresi wajah pada manusia menjadi isu yang begitu menarik untuk dikembangkan. Misalnya pada kamera pengawas pada sistem keamanan, bimbingan konseling bagi psikologi, menilai minat belajar siswa di kelas, me-monitoring sistem pengunjung dipusat perbelanjaan dan ditempat keramaian, serta mengenali statistik perilaku konsumen bagi industri retail. Tahapan dalam pengenalan ekspresi wajah manusia, terdiri dari dua tahapan, yang pertama adalah ekstraksi fitur dengan memperoleh ciri atau fitur yang penting pada wajah dari suatu citra gambar, selanjutnya adalah klasifikasi gambar yang merupakan pengklasifikasian citra pada kelas tertentu. Namun, dalam mendeteksi ekspresi wajah terdapat sejumlah tantangan yang perlu diperhatikan antara lain, posisi bentuk wajah, skala, ekspresi, wajah yang terhalangi objek lain, kondisi pencahayaan yang kurang[1], serta dibutuhkan data yang cukup banyak agar komputer dapat mengenali gambar tersebut. Selain itu, proses yang dibutuhkan cukup lama untuk mengumpulkan dan memproses gambar agar komputer dapat mengenali gambar tersebut. Oleh karena itu, peneliti tertarik untuk mengembangkan sistem pengenalan ekspresi wajah tanpa harus membutuhkan dataset yang banyak untuk melatih gambar. Dengan memanfaatkan metode *transfer learning* yaitu pemanfaatan model yang sudah dilatih oleh orang lain yang selanjutnya akan digunakan dalam mengenal ekspresi wajah pada manusia[2]. Adapun arsitektur yang digunakan adalah menggunakan *FaceNet*. *FaceNet* merupakan salah satu arsitektur *deep learning* yang digunakan untuk pengenalan, identifikasi, dan klasifikasi wajah yang selanjutnya dimodifikasi sehingga dapat mendeteksi ekspresi wajah.

II. LANDASAN TEORI

2.1. Pengenalan Ekspresi Wajah

Ekspresi wajah digunakan oleh manusia untuk menyampaikan berbagai jenis makna dalam berbagai situasi. Situasi dimana manusia merespons sebuah peristiwa di lingkungan ataupun sebagai bahasa isyarat untuk mengekspresikan suatu keadaan tertentu[3]. Mendeteksi emosi wajah dengan teknologi merupakan tugas yang cukup menantang namun, sangat menjanjikan di bidang *machine learning*. Pengenalan emosi sudah banyak digunakan oleh berbagai perusahaan antara lain untuk mengukur suasana hati konsumen terhadap produk atau merek mereka. Di bidang kesehatan dimanfaatkan untuk memutuskan kapan pasien memerlukan obat atau membantu dokter menentukan siapa yang harus dilihat terlebih dahulu, serta dapat dimanfaatkan pada pengenalan ekspresi wajah dalam pengujian video game yaitu membantu memahami emosi mana yang dialami pengguna secara real-time saat bermain tanpa menganalisis video lengkap secara manual[4].

2.2. *Transfer Learning*

Transfer Learning adalah pemanfaatan model yang diperoleh ketika memecahkan satu masalah dan menerapkannya pada masalah yang berbeda tetapi terkait[5]. Misalnya, pengetahuan yang diperoleh saat belajar mengenali mobil dapat digunakan sampai batas tertentu untuk mengenali becak atau truk.

2.3. *FaceNet*

FaceNet merupakan salah satu metode dari *deep convolutional network*. Metode ini disebut *one-shot learning* yaitu model yang dapat menggunakan sampel kecil gambar wajah untuk menghasilkan model awal, dan ketika ada model baru, model awal dapat digunakan tanpa melakukan pelatihan ulang[6]. *FaceNet* langsung melatih wajah menggunakan *uclidean space* dimana jarak terdiri dari kesamaan antara wajah dan model. Ketika hasil kesamaan antara model dan wajah diperoleh, maka akan mudah untuk melakukan pengenalan wajah dan klasifikasi menggunakan *FaceNet* menjadi *feature vectors*.

2.4. Computer Vision

Computer Vision adalah teknologi yang menggabungkan kamera, komputasi berbasis edge atau cloud, perangkat lunak, dan kecerdasan buatan (AI) untuk memungkinkan sistem “melihat” dan mengidentifikasi objek[7]. *Computer Vision* menggunakan pembelajaran mendalam dalam membentuk jaringan saraf (*neural network*) yang memandu sistem dalam pemrosesan dan analisis gambarnya. Jika sudah terlatih, maka model dari computer vision dapat melakukan pengenalan objek, mendeteksi dan mengenali orang hingga melacak pergerakan.

2.5. Pengolahan citra digital merupakan bidang ilmu informatika (Ilmu Komputer) yang mempelajari bagaimana suatu citra diolah, dibentuk, dan dianalisis sehingga akan menghasilkan informasi yang kemudian dapat dipahami oleh manusia[8]

2.6. Penelitian Terdahulu

Tahapan ini digunakan sebagai dasar dalam penyusunan kerangka berfikir penelitian. Selain itu, untuk mengetahui persamaan dan perbedaan dari beberapa penelitian terdahulu serta kajian yang dapat mengembangkan penelitian yang akan dilaksanakan. Berikut penyajian pustaka terdahulu dapat dirangkum dalam Tabel 1. sebagai berikut:

Tabel 1. Penelitian Terdahulu

No	Judul	Tahun, Penulis	Metode	Hasil/Kesimpulan
1	Face Recognition and Verification using Transfer Learning	2020, Sudhir Kumar, Athul AR, Amit Sethi	<i>Metode Teorema Bayes</i>	Tujuan dari penelitian ini adalah
2	Pengenalan Ekspresi Wajah Dengan Metode <i>Viola Jones</i> dan <i>Convolutional Neural Network</i>	2021, Reza Rizki Reynaldo, Irfan Maliki	<i>Metode Viola Jones dan Convolutional Neural Network</i>	Tujuan dari penelitian ini adalah untuk mengetahui dan meningkatkan akurasi pengenalan ekspresi wajah dengan menggunakan metode <i>Viola Jones</i> dan CNN. Metode yang dilakukan pada penelitian ini terdiri dari <i>preprocessing</i> , <i>segmentasi</i> menggunakan <i>Viola Jones</i> dan klasifikasi dengan CNN.
3	Pengenalan Ekspresi Wajah Dari <i>Cross Dataset</i> menggunakan <i>Convolutional Neural Network (CNN)</i>	2021, Prana Pramudita Kusdiananggalih, Ema Rachmawati, dan Risnandar	<i>Metode Convolutional Neural Network (CNN)</i>	Dataset yang digunakan adalah dengan dataset ras non-Indonesia dan dengan dataset ras Indonesia menggunakan teknik <i>cross dataset</i> . Hasil yang diperoleh dari penelitian ini adalah akurasi uji sebesar 91,29% <i>True Positive Rate (TPR)</i> sebesar 91,29%, <i>precision</i> sebesar 91,29%, dan <i>overall accuracy</i> sebesar 97,51%.

III. METODE

3.1. Dataset

Dataset yang digunakan pada penelitian ini adalah *Self Built Dataset*, yaitu dataset yang peneliti kumpulkan dengan mengambil foto pribadi menggunakan kamera dan opencv. Ekspresi dasar manusia pada penelitian ini terdiri dari ekspresi wajah normal, marah, dan senyum

dengan ukuran 640 X 480 pixels. Total keseluruhan data pada penelitian ini sejumlah 150 citra gambar dimana setiap masing-masing kelas terdiri dari 50 gambar. Berikut beberapa contoh *self build dataset* adalah sebagai berikut:

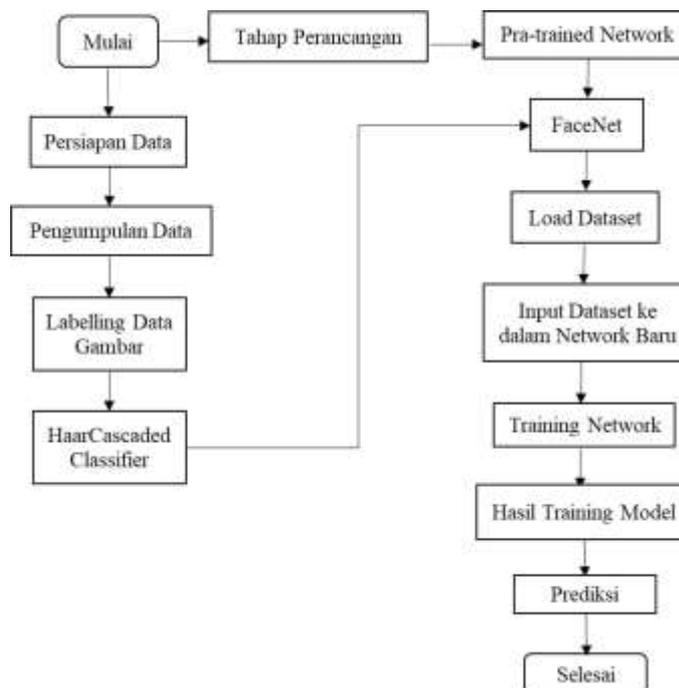


Gambar 1. Marah, Senang, Netral

Tabel 2. Definisi Operasional Variable Penelitian[9]

Variabel	Definisi Oprasional Variabel
Marah	Alis menyerong ke bawah Kelopak mata atas-bawah membuka lebar dan garis bibir melengkung ke bawah
Senang	Arah garis ekspresi wajah melengkung kearah atas, garis bibir membentuk lengkungan berbentuk U, dan bibir membuka lebar dengan memperlihatkan gigi
Netral	Gerakan bagian wajah datar

3.2. Perancangan Sistem



Gambar 2. Diagram *Flowchart* Rancangan Sistem

3.2.1. *Pra-trained Network*

Tahap perancangan pertama adalah menggunakan *pra-trained network* yang sudah dilatih sebelumnya untuk digunakan sebagai *Feature Extraction*. *Pra-trained* yang digunakan untuk mengekstrak *feature* dalam penelitian ini adalah dengan menggunakan *FaceNet*.

3.2.2 HaarCascaded Classifier

HaarCascaded Classifier digunakan untuk memotong wajah yang akan dideteksi kemudian data yang diperoleh akan diproses kedalam model *FaceNet*, agar menghasilkan output sebanyak 128 angka pada masing-masing gambar. Setelah mendapatkan hasil output dari *FaceNet* maka membuat network baru yang dapat digunakan untuk pendeteksian ekspresi wajah.

3.2.3. Pra-trained Network

Parameter masukan pada pra-proses *FaceNet* berupa ukuran inputan foto sebesar 160 X 160 pixels yang di kali 3(tiga) atau RGB. Ketika memberikan input kepada *FaceNet* maka *FaceNet* akan mentrain sehingga akan menghasilkan output berupa 128 *node* atau 128 angka.

128 angka ini yang mencirikan wajah yang telah di input. Jika wajah yang di input berbeda maka pada matriks 128 X 1 akan berbeda, jika menginput wajah dari orang yang sama meskipun berbeda pose maka angka yang dihasilkan mirip dengan angka yang sama sehingga disebut dengan "*signature*" atau sidik jari.

HaarCascaded Classifier berfungsi untuk mendeteksi berapa banyak wajah yang terdeteksi, Haarcascaded memiliki pola-pola tertentu yang disebut dengan *feature*[10]. Feature yang akan diujikan ke gambar yang ada, dan dari kecocokan gambar tersebut dengan feature-feature yang ada dan dari kecocokan gambar tersebut dengan feature-feature yang ada maka haarcascaded akan mengetahui mana yang termasuk foto wajah dan tidak.

Agar *FaceNet* dapat membaca gambar saat proses train maka perlu membuat signature dengan mengubah ukuran gambar menjadi dataset yang sebelumnya 640 X 480 pixels menjadi 160 X 160 pixels, kemudian mengubah gambar dari 3 dimensi menjadi 4 dimensi agar dapat dibaca ketika menjalankan model. Ketika memasukkan dataset ke haarcascaded maka akan menghasilkan input berupa deteksi wajah dan setelah diolah pada haarcascaded maka *FaceNet* akan menghasilkan signature.

3.2.4. Load Dataset

Tahap selanjutnya yaitu meload dataset yang telah dibuat sebelumnya, fungsi dari load data yaitu agar data yang telah dibuat dapat diproses untuk menghasilkan signature. Hasil dari load dataset yaitu, trainX sama dengan 150 X 128 dan trainY sebesar 150 X 1. 150 merupakan keseluruhan gambar yang telah diproses, sedangkan 128 merupakan output angka dari *FaceNet*.

3.2.5. Input Dataset ke dalam Network dengan Layer Baru

Selanjutnya adalah menginputkan dataset ke dalam network yang telah dibuat dengan beberapa layer. Layer pertama adalah 128 karena sebelumnya adalah 128 angka menggunakan *relu* dan terdiri dari 20 *node*, layer ke 2 dan ketiga terdiri dari 10 *node* dan *activation relu*. Layer terakhir terdiri dari 3 *node* dan untuk *activation* nya yaitu *softmax*.

3.2.6. Training Network

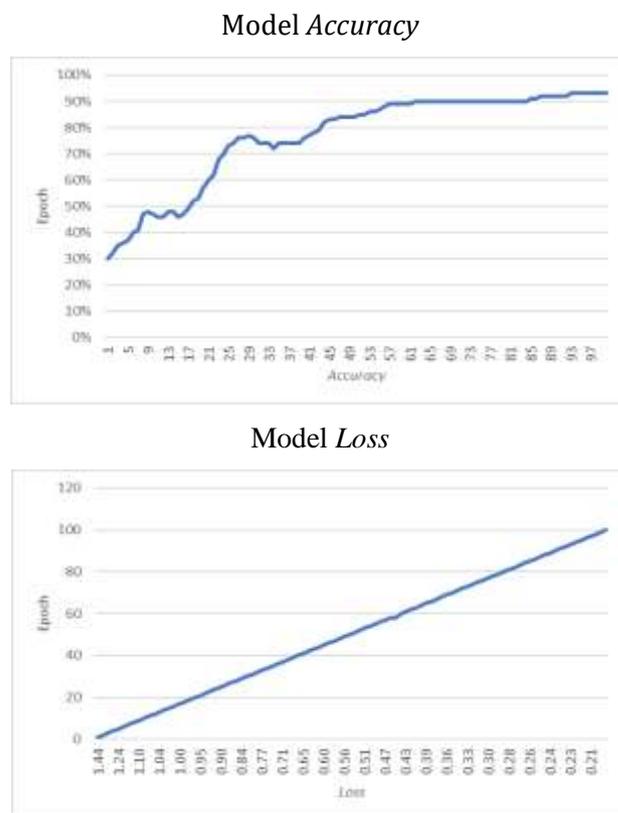
Setelah mendapatkan dataset dan membuat layer maka berikutnya adalah mentrain network yang telah dibuat agar dapat membedakan 3 ekspresi wajah yaitu, marah, netral, dan senang. Epochs nya 100 sedangkan batch size nya adalah 150, karena terdiri dari 150 gambar per input. Semua percobaan diatas dilatih menggunakan GPU laptop agar proses *training* dapat berjalan lebih cepat. Akurasi yang didapatkan dari hasil *training* yaitu sebesar 94%.

3.2.7. Hasil Training Model

Model hasil *training* merupakan model yang akan digunakan dalam mendeteksi gambar. Model ini berekstensi (.h5) dimana pada model ini terdapat *weight* (bobot), label semua kelas dan arsitektur *deep learning* model *FaceNet* secara keseluruhan.

IV. HASIL DAN PEMBAHASAN

Hasil penelitian dan pembahasan pada penelitian ini menjelaskan hasil dari uji coba model yang telah dirancang dengan tujuan untuk mengetahui apakah model yang dirancang dapat berjalan dengan baik sebagaimana yang diinginkan. Pada penelitian ini, hasil pengujian diujikan pada proses pengenalan serta prediksi emosi seseorang secara real time. Pengujian dilakukan dengan *epoch* 100 dan *batch size* senilai 150. Hasil dari pengujian dapat dilihat pada **Gambar 3**:



Gambar 3. Grafik Accuracy dan Loss

4.1. Pengujian kemampuan model transfer learning dan *FaceNet* dalam mengenali dan mendeteksi wajah (recognize)

Pada penelitian ini pengujian model transfer learning pada *FaceNet* menggunakan webcam dengan posisi wajah menghadap kamera. Pada percobaan ini model transfer learning pada *FaceNet* berhasil mengenali tiga klasifikasi wajah yaitu, wajah senang, marah, dan netral secara realtime. Untuk pengujian dapat dilihat pada **Gambar 4**.



Gambar 4. Hasil Uji Kemampuan Model Transfer Learning pada *FaceNet*

V. KESIMPULAN

Berdasarkan penelitian ini maka, dapat disimpulkan bahwa jumlah dataset yang sedikit tidak mempengaruhi hasil akurasi. Dengan kemampuan model transfer learning peneliti dapat mempelajari pengetahuan baru yaitu algoritma *FaceNet* yang sebelumnya dimanfaatkan dalam pengenalan dan klasifikasi wajah dengan penelitian ini *FaceNet* dapat dimanfaatkan untuk mendeteksi ekspresi wajah yaitu ekspresi wajah senang, marah, dan netral.

Metode *FaceNet* sangat cocok digunakan untuk menguji sebuah citra karena *FaceNet* langsung melatih wajah menggunakan *uclidean space* dimana jarak terdiri dari kesamaan antara wajah dan model. Sehingga ketika terjadi kesamaan antara wajah dan model maka *FaceNet* dapat mendeteksi wajah secara langsung.

Penentuan *Batch size* dan *Epoch* sangat mempengaruhi akurasi yang akan didapatkan. Dengan *Epoch 100*, *Batch size 150* didapatkan hasil akurasi *training* sebesar 93%

DAFTAR PUSTAKA

- [1] X. Yang, S., Luo, P., Loy, C. C., & Tang, "WIDER FACE: A Face Detection Benchmark," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, no. 1063–6919, pp. 5525–5533, 2016, [Online]. Available: <https://ieeexplore.ieee.org/document/7780965>.
- [2] R. Harish, "Transfer And Ye Shall Learn: An overview of Transfer Learning," *medium*, 2021. <https://medium.com/sfu-csmpmp/transfer-and-ye-shall-learn-an-overview-of-transfer-learning-a118855a3ebd>.
- [3] E. A. Elliott and A. M. Jacobs, "Facial expressions, emotions, and sign languages," *Front. Psychol.*, vol. 4, no. MAR, pp. 4–7, 2013, doi: 10.3389/fpsyg.2013.00115.
- [4] sightcorp, "Everything About EMOTION RECOGNITION," <https://sightcorp.com/>. <https://sightcorp.com/knowledge-base/emotion-recognition/>.
- [5] A. Ar and A. Sethi, "Face Recognition and Verification using Transfer Learning," no. November, 2020, doi: 10.13140/RG.2.2.26851.99367.
- [6] J. Williams, "Deep Learning In Digital Pathology," *global-engage*, 2018. <http://www.global-engage.com/life-science/deep-learning-in-digital-pathology/> (accessed Dec. 10, 2021).
- [7] K. Pulli, A. Baksheev, K. Korniyakov, and V. Eruhimov, "Realtime computer vision with OpenCV," *Queue*, vol. 10, no. 4, pp. 1–17, 2012, doi: 10.1145/2181796.2206309.
- [8] W3schools, "Python Introduction." https://www.w3schools.com/python/python_intro.asp (accessed Jun. 26, 2021).
- [9] K. Amda and R. Fitriyani, "Membaca Ekspresi Wajah Manusia." Huta Publisher, Depok, p. 206, 2016.
- [10] Aditya Mittal, "Haar Cascades, Explained," *medium.com*, 2020. <https://medium.com/analytics-vidhya/haar-cascades-explained-38210e57970d>.