

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengenalan ekspresi wajah telah menjadi salah satu fondasi utama dalam bidang visi komputer saat ini, dengan peran krusial dalam pengembangan interaksi manusia-komputer (*Human-Computer Interaction*) yang lebih natural dan responsif [5]. Teknologi ini memungkinkan sistem komputasi untuk secara otomatis memahami keadaan emosional penggunanya, membuka peluang aplikasi di berbagai sektor seperti sistem pengawasan pengemudi pintar untuk keamanan berkendara, hingga analisis sentimen perilaku konsumen dalam studi pasar [4]. Sejalan dengan meningkatnya permintaan terhadap sistem yang cerdas dan adaptif, kebutuhan akan akurasi model pengenalan ekspresi wajah yang robust semakin mendesak, khususnya dalam kondisi *real-world* yang kompleks [10].

Namun, penerapan sistem pengenalan ekspresi wajah dalam kondisi nyata (*in-the-wild*) menghadapi tantangan yang jauh lebih rumit dibandingkan lingkungan laboratorium yang terkontrol. Dataset standar industri seperti FER-2013 menggambarkan kompleksitas ini secara efektif, menyajikan variasi data yang ekstrem [1]. Gambar wajah dalam dataset ini sering kali memiliki resolusi rendah, penerangan yang tidak konsisten, banyak gangguan visual (*noise*), serta variasi pose wajah yang tidak frontal. Selain itu, keberadaan oklusi atau bagian wajah yang tertutup (misalnya oleh tangan, kacamata, atau aksesoris) semakin mempersulit ekstraksi fitur wajah yang lengkap dan representatif [11].

Tantangan teknis yang lebih mendasar terletak pada kemiripan visual (*inter-class similarity*) yang tinggi antara kelas emosi tertentu. Sebagai contoh, ekspresi 'Takut' (*Fear*) dan 'Terkejut' (*Surprise*) sering menunjukkan pola perubahan wajah yang sangat serupa, di mana perbedaan utamanya hanya terdapat pada rincian mikro seperti kekuatan buka mata atau tarikan sudut bibir [1]. Arsitektur *Convolutional Neural Network* (CNN) konvensional sering kali mengalami kesulitan dalam menangkap nuansa perbedaan halus semacam ini, karena operasi konvolusi dan *pooling* standar cenderung memperlakukan seluruh area *feature map* dengan

bobot kepentingan yang sama, tanpa secara selektif menekankan region-region kritis yang mengandung informasi emosional paling informatif [20].

Masalah berikut yang kerap muncul dalam pelatihan *Deep Learning* pada dataset yang terbatas adalah risiko *overfitting* dan kepercayaan berlebihan. Dataset FER-2013 mengalami masalah ketidakseimbangan jumlah data antar kelas yang signifikan, di mana kelas '*Happy*' memiliki jumlah sampel jauh lebih banyak daripada kelas '*Disgust*'. Pemanfaatan model yang sangat mendalam pada situasi data yang tidak seimbang ini sering mengakibatkan model condong kepada kelas mayoritas. Di samping itu, penerapan label target "keras" dalam fungsi kehilangan standar cenderung memaksa model untuk terlalu percaya diri pada prediksinya, sehingga menurunkan kemampuan generalisasi model ketika menghadapi data uji yang memiliki ambiguitas ekspresi [2]

Untuk mengatasi kekurangan data dan kompleksitas pelatihan dari awal, pendekatan *Transfer Learning* menjadi solusi yang sangat tepat. Penelitian ini menggunakan kerangka kerja *DeepFace* untuk menerapkan arsitektur *VGG-16* sebagai pengekstraksi fitur utama. *VGG-16* dipilih karena struktur arsitekturnya yang dalam dengan 13 lapisan konvolusi serta kemampuannya yang terbukti handal dalam mempelajari representasi fitur hierarkis wajah dari dataset besar (*VGG-Face*). Pemanfaatan framework *DeepFace* memfasilitasi standarisasi pra-pemrosesan data, menjamin bahwa model *VGG-16* memperoleh data wajah yang telah teridentifikasi dan diselaraskan dengan tepat.

Sebagai inovasi utama untuk menangani isu "kemiripan visual" yang telah disebutkan, studi ini mengusulkan penerapan mekanisme atensi *Squeeze-and-Excitation (SE-Block)* dalam arsitektur *VGG-16*. Modul *SE-Block* berfungsi dengan melakukan kalibrasi ulang bobot fitur secara adaptif. Dengan cerdas, modul ini akan memberikan perhatian lebih pada channel fitur yang menggambarkan area penting (seperti mata dan mulut) dan menurunkan (*suppress*) fitur yang kurang relevan atau kebisingan. Diharapkan bahwa hal ini dapat meningkatkan kepekaan model dalam membedakan ekspresi mikro yang susah.

Selain inovasi dalam arsitektur, penelitian ini juga menggunakan strategi optimasi pada fase pelatihan dengan metode *Label Smoothing*. Metode regularisasi

ini bertujuan untuk menghaluskan label target, menghindari model menjadi terlalu yakin (*overconfident*) terhadap satu kategori tertentu, serta mendorong terbentuknya kluster fitur yang lebih solid dan jelas terpisah. Dengan cara ini, model diharapkan dapat lebih toleran terhadap kesalahan pelabelan (*noise label*) yang mungkin ada dalam dataset dan lebih tangguh terhadap overfitting dibandingkan metode pelatihan tradisional.

Terakhir, untuk meningkatkan ketepatan prediksi selama fase pengujian, penelitian ini menggunakan metode *Test Time Augmentation* (TTA). Daripada hanya bergantung pada satu prediksi deterministik, TTA memungkinkan model untuk melakukan inferensi pada berbagai variasi augmentasi dari gambar uji yang serupa (misalnya, versi yang terbalik atau sedikit bergeser). Hasil analisis dari berbagai "perspektif" ini kemudian dirata-ratakan untuk memperoleh keputusan akhir yang lebih konsisten. Penggabungan menyeluruh antara *backbone* VGG-16 yang tangguh, mekanisme atensi *SE-Block*, serta pendekatan *Label Smoothing* dan TTA adalah dasar utama penelitian untuk menciptakan sistem klasifikasi ekspresi wajah yang terbaik.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Memiliki banyak tantangan untuk mengidentifikasi ekspresi wajah secara otomatis, terutama terkait dengan variasi ekspresi, pencahayaan, sudut pandang, dan latar belakang. Metode konvensional seringkali tidak cukup efektif untuk menangani perubahan emosi yang kompleks yang terjadi pada foto wajah. Selain itu, dataset ekspresi wajah seperti FER2013 menghadapi beberapa masalah. Ini termasuk ketidakseimbangan kelas, resolusi yang rendah (48 x 48 piksel), dan keterbatasan untuk menangkap variasi kondisi pencahayaan dan pose wajah. Akibatnya, model kesulitan membedakan ekspresi yang serupa, seperti ekspresi netral dan sedih [3].

1.2.2 Solusi Permasalahan

Untuk mengatasi masalah diatas, penelitian ini menyarankan penggunaan *transfer learning* berbasis *DeepFace* yang terintegrasi CNN. *DeepFace* berfungsi sebagai *feature extractor*, yang memungkinkan model menggunakan fitur ekspresi

wajah yang telah dipelajari dari dataset yang sangat besar. Selanjutnya, untuk meningkatkan kinerja klasifikasi. Untuk meningkatkan kualitas data FER2013, teknik *preprocessing* seperti *face detection*, *resizing*, normalisasi, dan augmentasi yang diperluas. Ini memungkinkan model untuk mendeteksi ekspresi wajah dengan lebih baik.

1.2.3 Pertanyaan Penelitian

- 1) Bagaimana kinerja transfer learning berbasis *DeepFace* dalam klasifikasi ekspresi wajah menggunakan dataset FER2013?
- 2) Bagaimana kinerja *DeepFace* dalam klasifikasi ekspresi wajah pada dataset FER2013?
- 3) Bagaimana kinerja dan performa SE-Block dalam klasifikasi ekspresi wajah pada dataset FER2013?

1.3 Batasan Masalah

Agar penelitian ini lebih terfokus, beberapa batasan yang diterapkan dalam penelitian ini adalah sebagai berikut:

- 1) Dataset yang digunakan adalah FER-2013 (*Facial Expression Recognition* 2013) yang terdiri dari 7 kelas emosi: *Angry*, *Disgust*, *Fear*, *Happy*, *Sad*, *Surprise*, dan *Neutral*.
- 2) Arsitektur model utama yang digunakan adalah *VGG-16* yang dimuat melalui library *DeepFace*, dengan bobot awal (*pre-trained*) dari *VGG-Face*.
- 3) Modifikasi arsitektur difokuskan pada penambahan modul SE-Block setelah backbone konvolusi terakhir, serta penggunaan teknik Label Smoothing pada fungsi kerugian (loss function).
- 4) Evaluasi kinerja model diukur menggunakan metrik Akurasi, Presisi, Recall, F1-Score, dan analisis Confusion Matrix.
- 5) Penelitian ini tidak membahas klasifikasi identitas, jenis kelamin, atau estimasi umur, melainkan fokus sepenuhnya pada klasifikasi 7 jenis ekspresi wajah.

- 6) Pelatihan model dilakukan dengan spesifikasi perangkat terbatas, yaitu prosesor Intel Core i5 Gen 11 dengan NVIDIA MX350 dan RAM 12GB.

1.4 Tujuan dan Manfaat

1.4.1 Tujuan

- 1 Mengimplementasikan arsitektur Transfer Learning VGG-16 yang dimodifikasi dengan *Squeeze-and-Excitation (SE-Block)* untuk klasifikasi ekspresi wajah.
- 2 Menganalisis perbandingan kinerja antara Skenario 1 (Model *Baseline/Frozen*) dan Skenario 2 (Model *Optimized* dengan *SE-Block*, *Label Smoothing*, dan TTA) dalam mengenali ekspresi wajah pada dataset FER-2013.

1.4.2 Manfaat

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Bagi Akademisi: Memberikan kontribusi wawasan mengenai efektivitas integrasi mekanisme atensi (*Attention Mechanism*) dan teknik regularisasi modern pada arsitektur VGG-16 untuk tugas pengenalan emosi.
2. Bagi Pengembang Teknologi: Menjadi referensi teknis dalam membangun sistem pengenalan wajah yang lebih *robust* terhadap variasi data *in-the-wild*.
3. Bagi Masyarakat: Potensi penerapan dalam sistem deteksi emosi untuk evaluasi kepuasan pelanggan atau pemantauan kondisi psikologis secara otomatis.

1.5 Sistematika Laporan

Sistematika laporan dalam penelitian ini disusun sebagai berikut:

- **BAB I Pendahuluan**

Bab ini membahas latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat, serta sistematika laporan.

- **BAB II Kajian Pustaka**

Bab ini menjelaskan teori-teori yang mendukung penelitian, termasuk konsep dasar CNN, klasifikasi ekspresi wajah, *transfer learning*, *Deepface*, serta penelitian terdahulu yang relevan.

- **BAB III Metodologi Penelitian**

Bab ini berisi metode yang digunakan dalam penelitian, termasuk pengumpulan data, preprocessing, arsitektur model, dan teknik evaluasi.