

1. Pendahuluan

Latar Belakang

Pemrosesan wajah (*face processing*) merupakan salah satu teknologi dalam bidang *computer vision* yang paling berkembang saat ini. Salah satunya yaitu Pengenalan Ekspresi Wajah (*Facial Expression Recognition*). Pengenalan ekspresi wajah sering dikaitkan dengan emosi. Pada bidang marketing, kepuasan pelanggan terhadap pelayanan penting untuk diperhatikan oleh perusahaan. Salah satu indikator yang dapat menunjukkan kepuasan pelanggan terhadap pelayanan adalah ekspresi wajah atau emosi pelanggan. Dengan adanya Pengenalan Ekspresi Wajah, perusahaan telah dibukakan peluang untuk meningkatkan tingkat loyalitas pelanggan terhadap perusahaan. Oleh, karena itu penulis tertarik untuk membangun sistem pengenalan ekspresi wajah.

Sejauh ini, sudah banyak peneliti yang melakukan penelitian mengenai Pengenalan Ekspresi Wajah (*Facial Expression Recognition*) hingga tahun 2019 [1]. Dari pendekatan konvensional hingga pendekatan *Deep Learning*. *Dataset* yang digunakan juga bermacam-macam antara lain JAFFE, CK+, MMI, FERA, FER2013, BU-3DFE, Multi-PIE, SFEW, Oulu-Casia. Salah satu penelitian menggunakan cara konvensional yaitu metode Gabor + *Support Vector Machine* (SVM) diterapkan pada *dataset* JAFFE menghasilkan tingkat akurasi sebesar 80,95% [1], [2]. Untuk salah satu penelitian menggunakan *Deep Learning* salah satunya yaitu CNN, pada *dataset* FER2013 hanya mencapai akurasi sebesar 72,10% [1],[3]. Lalu ada dua penelitian yang menggunakan *dataset* yang sama yaitu FER2013 menggunakan metode CNN (*DeeperCNN*) dan *CubicSVM+HOG* mendapat akurasi sebesar 61,10% dan 57,17% [1]. Sedangkan menggunakan metode yang sama yaitu CNN pada *dataset* lain antara lain, CK+ mencapai 98,62%, MMI mencapai 63,40%, FERA mencapai 76,70%, Multi-PIE mencapai 94,70%, SFEW mencapai 47,70%, dan Oulu-CASIA mencapai 58,18% [1]. Hasil akurasi pada tiap *dataset* dan masing-masing metode yang digunakan memiliki akurasi yang berbeda-beda, tentunya banyak sekali parameter-parameter atau faktor-faktor yang dapat mempengaruhi tingkat akurasi. Salah satunya yaitu masing-masing *dataset* memiliki tantangannya tersendiri.

Hampir seluruh penelitian yang dilakukan saat ini, menggunakan *dataset* yang berisi data orang-orang non-Indonesia sebagai data latih maupun data uji [1]. Hal ini memberikan peneliti ide untuk menerapkan metode *Cross Dataset* di mana data latih dan data uji menggunakan *dataset* yang berbeda [3], FER2013 sebagai data latih dan Data-data orang Indonesia (yang peneliti kumpulkan sendiri) sebagai data uji.

Pada suatu penelitian yang dilakukan pada tahun 2016 oleh Mollahosseini, Ali Chan, David Mahoor, dan Mohammad H [4]. Telah dilakukan uji coba *proposed architecture CNN*, yaitu membangun arsitektur CNN sendiri untuk dilakukan uji coba terhadap beberapa *dataset* salah satunya terdapat FER2013 [4]. Pada penelitian tersebut, dilakukan perbandingan antara *proposed architecture CNN* dengan arsitektur populer yaitu AlexNet. Untuk *dataset* FER2013 menggunakan *proposed architecture CNN* menghasilkan akurasi sebesar 66,4%, sedangkan AlexNet menghasilkan akurasi sebesar 61,1% [4]. Hasil lain pada *dataset* MMI dihasilkan akurasi sebesar 77,9% pada *proposed architecture CNN*, sedangkan AlexNet menghasilkan akurasi sebesar 56,0% [4]. Hal ini menunjukkan bahwa arsitektur yang dibangun sendiri memungkinkan lebih baik daripada *DeeperCNN* seperti AlexNet untuk beberapa *dataset*.

Pengenalan ekspresi wajah sudah banyak dilakukan penelitiannya sampai tahun 2019 [1], akan tetapi untuk penerapan pengenalan ekspresi wajah menggunakan *dataset* orang-orang Indonesia, masih sedikit. Oleh karena itu, peneliti mengusulkan untuk membangun sistem pengenalan ekspresi wajah secara *Cross Dataset* menggunakan metode *Convolutional Neural Network* (CNN).

Topik dan Batasannya

Definisi beberapa dari kata kunci yang ada pada abstrak :

1. Facial Expression Recognition (FER)
Atau dalam Bahasa Indonesia artinya pengenalan ekspresi wajah manusia adalah sistem yang dapat mengklasifikasi sebuah citra wajah manusia kedalam kategori emosi seperti bahagia, netral, sedih, marah, terkejut, jijik, dan takut.
2. Convolutional Neural Network (CNN)
Merupakan metode *Deep Learning* yang sangat populer dan banyak digunakan hingga saat ini. Metode ini sangat populer karena kecepatannya dalam melakukan *training data*, tersedia banyak library (contoh : *KERAS*), tidak perlu membuat fitur ekstraksi secara manual
3. *Cross Dataset*
Cross dataset adalah melakukan silang antara data latih dan data uji antar *dataset*. Sebagai contoh dataset CK+, FER2013 dan NovaEmotions. Masing-masing *dataset* memiliki data latih dan data uji. Dan masing-masing data latih dan data uji akan dilakukan silang, sehingga menghasilkan tingkat akurasinya masing-masing. Berikut merupakan contoh *cross dataset* yang digunakan dapat dilihat pada tabel 1.1 :

Tabel 1.1 Contoh tabel *cross dataset* [3]

Train \ Test	CK+	FER2013	NovaEmotions
CK+	98.62%	69.3%	67.2%
FER2013	92.0%	72.1%	78.0%
NovaEmotions	93.75%	71.8%	81.3%

Berikut terdapat beberapa Batasan-batasan masalah :

1. *Dataset* yang digunakan adalah FER2013 dan *Dataset* yang dibangun sendiri
2. *Dataset* yang dibangun sendiri diberi nama “*Self built dataset*”
3. *Dataset* yang dibangun sendiri diambil dari internet lalu dikumpulkan dan dilabeli secara manual
4. *Dataset* FER-2013 yang digunakan berjumlah 35.685 (sebagai data latih dan data validasi)
Jumlah *dataset* FER-2013 secara keseluruhan yaitu 35.685, namun yang akan digunakan sebagai data latih berjumlah 28.709 dan sisanya berjumlah 6.976 digunakan sebagai data validasi
5. Kedua *dataset* yang digunakan berisi data dengan informasi piksel (0-255) dengan ukuran matriks 48x48 (bukan data berbentuk format citra JPEG, PNG dsb.) yang disimpan dalam format CSV
6. *Dataset* yang dibangun sendiri berjumlah terbatas, yaitu 379 (sebagai data uji)
Karena keterbatasan sumber daya dan waktu peneliti hanya sanggup mengumpulkan kurang lebih 400 citra berisi ekspresi wajah orang Indonesia sebagai data uji. Proses pengumpulan data dilakukan secara manual yaitu dibagi menjadi berbagai tahapan antara lain pengumpulan referensi media yaitu foto wajah orang Indonesia di Internet, Youtube, Film. Lalu dilakukan *crop scene/frame* dengan ukuran 48x48 (1:1 rasio) berisikan ekspresi wajah orang Indonesia secara manual. Setelah itu disimpan dalam format JPEG, lalu dimasukkan ke dalam masing-masing folder kategori emosi secara manual (sebagai labeling). Peneliti membuat code python yang akan melakukan ekstraksi citra format JPEG menjadi sebuah barisan angka dalam format 0-255 dan akan memberikan label berdasarkan letak file citra dalam folder yang telah dilabel secara manual lalu disimpan ke dalam format CSV. Selain itu, terdapat beberapa ekspresi wajah yang sulit didapat yaitu ekspresi ‘jijik’ atau *disgust* sangat sulit didapatkan di Internet, Film, media. Oleh karena itu, jumlah *dataset* ‘jijik’ atau *disgust* sangat terbatas.
7. *Dataset* memiliki format warna abu-abu (*Grayscale*)
Dataset FER-2013 memang memiliki format *grayscale*, oleh karena itu, pada *dataset* yang dibangun sendiri oleh peneliti, harus dikonversi dari RGB menjadi *grayscale* menggunakan metode *weighted method*
8. *Dataset* memiliki ukuran 48x48 yaitu rasio 1:1
Sehingga citra gambar hanya merupakan citra wajah, sehingga tidak perlu melakukan *Face Detection* pada penelitian ini
9. Ekspresi wajah dikategorikan menjadi tujuh klasifikasi
Yaitu antara lain bahagia, netral, sedih, marah, terkejut, jijik, dan takut
10. *Dataset* FER2013 memiliki banyak anomali
Berdasarkan penelitian FER pada *dataset* FER2013, akurasi yang didapat sekitar 57.17-61.10% hal ini dikarenakan *dataset* memiliki tantangan kondisi tersendiri. Subjek memiliki ekspresi spontan yang di bawah berada pada keadaan yang unik. Bahkan terdapat wajah citra kartun pada *dataset*. [1]
11. Batas usia dan jenis kelamin objek manusia
Untuk *dataset* FER2013 terdapat beragam usia dan jenis kelamin manusia antara lain usia bayi (0-1 tahun), anak-anak (2-10 tahun), remaja (11-19 tahun), dewasa (20-60 tahun) serta jenis kelamin laki-laki dan perempuan. Untuk *self built dataset* antara lain anak-anak (2-10 tahun), remaja (11-19 tahun), dewasa (20-60 tahun) dengan jenis kelamin laki-laki dan perempuan.

Tujuan

Tujuan dari Tugas Akhir (TA) ini adalah membangun sistem pengenalan ekspresi wajah dari *cross dataset* menggunakan metode *Convolutional Neural Network* (CNN).

Organisasi Tulisan

Laporan tugas akhir ini terdiri dari lima bagian. Pada bagian pertama menjelaskan pendahuluan, pada bagian kedua menjelaskan studi terkait, pada bagian ketiga menjelaskan sistem yang dibangun, pada bagian keempat menjelaskan evaluasi dari sistem yang dibangun, dan pada bagian kelima menjelaskan kesimpulan yang diambil dari tugas akhir ini.