

BAB I

PENDAHULUAN

Bab ini menyajikan pendahuluan dari penelitian yang dilakukan. Pembahasan mencakup latar belakang yang menjadi dasar dilakukannya penelitian, perumusan masalah yang ingin diselesaikan, serta tujuan dan manfaat penelitian baik secara teoritis maupun praktis. Selain itu, bab ini juga memuat batasan penelitian yang menjelaskan ruang lingkup studi agar tetap fokus dan terarah, serta gambaran struktur organisasi penulisan skripsi secara keseluruhan.

1.1. Latar Belakang Penelitian

Ekspresi wajah merupakan salah satu bentuk komunikasi non-verbal yang penting dalam interaksi manusia (Frith, 2009). Dalam konteks pembelajaran di kelas, ekspresi wajah dapat mencerminkan berbagai keadaan emosional manusia, seperti kemarahan, kebahagiaan, ketakutan, keterkejutan, rasa jijik, dan kesedihan, yang sering kali tidak diungkapkan secara verbal (Batty & Taylor, 2003). Pemahaman terhadap ekspresi wajah ini dapat membantu pendidik mengidentifikasi kebutuhan emosional dan akademik peserta didik, sehingga memungkinkan mereka memberikan dukungan yang lebih tepat dan efektif (Julika & Setiyawati, 2019).

Seiring dengan perkembangan teknologi, pengenalan ekspresi wajah atau *Facial Expression Recognition* (FER) telah berkembang menjadi sistem komputer yang mampu menganalisis dan mengenali gerakan wajah serta perubahan fitur wajah dari informasi visual (Tian et al., 2011). *Facial Expression Recognition* (FER) memiliki berbagai manfaat dalam berbagai bidang kehidupan, termasuk interaksi manusia-komputer, realitas virtual, sistem bantuan pengemudi canggih, hiburan, dan pendidikan (Huang et al., 2019). Dalam konteks pendidikan, FER berperan penting untuk memahami emosi siswa melalui ekspresi wajah, yang dapat memberikan informasi berharga bagi pendidik. Studi menunjukkan bahwa emosi seperti keterlibatan dan frustrasi memengaruhi hasil pembelajaran secara signifikan, dengan ekspresi tertentu, seperti *mouth dimpling*, diidentifikasi sebagai indikator positif peningkatan pembelajaran (Grafsgaard et al., 2013). Pemantauan

ekspresi wajah siswa memungkinkan pendidik untuk menyesuaikan metode pengajaran mereka agar lebih efektif (Hikmatiar et al., 2024).

FER juga menjadi salah satu topik utama penelitian dalam bidang kecerdasan buatan (*artificial intelligence* - AI) dan visi komputer (*computer vision*), karena kemampuannya untuk menganalisis dan mengenali emosi manusia dari informasi visual (Canedo & Neves, 2019). Teknologi berbasis *machine learning*, seperti Convolutional Neural Networks (CNN), telah digunakan dalam berbagai aplikasi, termasuk pengenalan objek, klasifikasi gambar, segmentasi gambar, pemrosesan video, pemrosesan bahasa alami, dan pengenalan suara (Bhatt et al., 2021). CNN telah banyak digunakan untuk berbagai aplikasi, termasuk keamanan (Sunardi et al., 2022), ketertiban lalu lintas (Nurhawanti, 2019), deteksi kendaraan (Yudha, 2020), kesehatan (Febriyanti, 2024), pertanian (A., 2023), peternakan (Wulandari & Jasmir, 2024), dan pendidikan (Anggraini & Zakaria, 2023).

Dalam tugas pengenalan ekspresi wajah, CNN memiliki keunggulan dalam mengekstraksi fitur penting dari gambar secara otomatis, yang membuatnya sangat efektif untuk tugas pengenalan visual (Moshayedi et al., 2022). CNN menunjukkan akurasi tinggi pada emosi tertentu seperti "happy" dan "surprise," namun menghadapi tantangan signifikan, termasuk sensitivitas terhadap pencahayaan, posisi kamera, dan kesulitan mendeteksi ekspresi mikro (Fakhar et al., 2022).

Sebagai alternatif yang menjanjikan, Vision Transformer (ViT) diperkenalkan oleh (Dosovitskiy et al., 2020) sebagai model *machine learning* yang dirancang untuk tugas klasifikasi gambar menggunakan arsitektur Transformer. Berbeda dengan CNN, ViT menggunakan mekanisme perhatian (*attention mechanism*) yang mampu menangkap hubungan antar-piksel secara lebih kontekstual dengan cara memproses gambar melalui pembagian menjadi serangkaian *patch*, yang kemudian dianalisis untuk mengenali pola visual dengan akurasi tinggi (Han et al., 2020).

Di Indonesia, penelitian terkait pengenalan ekspresi wajah di lingkungan pendidikan masih tergolong minim, meskipun teknologi seperti ViT memiliki potensi besar dalam membantu pendidik memantau kondisi emosional peserta didik secara real-time. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem pengenalan ekspresi wajah berbasis ViT yang tidak hanya memberikan

akurasi tinggi, tetapi juga mampu menangani variasi ekspresi wajah dalam skenario ruang kelas. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi pendidikan berbasis kecerdasan buatan serta memperkaya literatur ilmiah terkait penggunaan ViT dalam bidang visi komputer.

1.2. Rumusan Masalah Penelitian

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengembangkan model pengenalan ekspresi wajah berbasis Vision Transformer (ViT) untuk digunakan dalam lingkungan kelas?
2. Bagaimana perbandingan efektivitas model berbasis Vision Transformer (ViT) dengan metode lain seperti YOLO (You Only Look Once) dalam mengenali berbagai macam ekspresi wajah?
3. Bagaimana pengaruh berbagai faktor terhadap performa model Vision Transformer (ViT) dalam tugas pengenalan ekspresi wajah peserta didik di lingkungan kelas?
4. Bagaimana performa keseluruhan sistem pengenalan ekspresi wajah berbasis Vision Transformer (ViT) dalam ruang kelas, dan apakah sistem tersebut dapat dimanfaatkan untuk membantu pendidik memantau emosi peserta didik secara real-time?

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan model pengenalan ekspresi wajah berbasis Vision Transformer (ViT) yang dirancang khusus untuk lingkungan kelas.
2. Mengevaluasi efektivitas model berbasis Vision Transformer (ViT) dalam mengenali berbagai macam ekspresi wajah dibandingkan dengan metode lain seperti YOLO (You Only Look Once).
3. Mengidentifikasi pengaruh berbagai faktor terhadap performa model Vision Transformer (ViT) dalam tugas pengenalan ekspresi wajah peserta didik di lingkungan kelas.

4. Menilai performa keseluruhan sistem pengenalan ekspresi wajah berbasis Vision Transformer (ViT) dalam konteks ruang kelas, serta mengevaluasi potensi pemanfaatannya dalam membantu pendidik memantau kondisi emosional peserta didik secara real-time.

1.4. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan ilmu pengetahuan, khususnya di bidang kecerdasan buatan, visi komputer, dan pendidikan, di antaranya:

1. Pengayaan literatur terkait penggunaan Vision Transformer (ViT) dalam pengenalan ekspresi wajah. Hasil penelitian ini dapat memperluas pengetahuan tentang kemampuan ViT dalam menangani data visual, terutama dalam konteks variasi ekspresi wajah yang subtil atau kompleks.
2. Pengembangan model dan teknik pengenalan ekspresi wajah dalam lingkungan pendidikan. Penelitian ini akan memperkenalkan metode pengenalan ekspresi wajah di ruang kelas yang bisa menjadi dasar bagi penelitian selanjutnya di bidang yang serupa.
3. Pemahaman lebih lanjut tentang faktor-faktor yang mempengaruhi akurasi pengenalan ekspresi wajah. Penelitian ini akan mengidentifikasi faktor-faktor kunci yang dapat berperan dalam meningkatkan atau mengurangi kinerja sistem pengenalan wajah berbasis ViT.

1.5. Batasan Penelitian

Untuk menjaga fokus penelitian dan memudahkan dalam pengolahan data serta analisis, penelitian ini memiliki beberapa batasan sebagai berikut:

1. Lingkup Ekspresi Wajah yang Dikenali

Penelitian ini hanya mengenali tujuh jenis ekspresi wajah dasar, yaitu bahagia, sedih, marah, takut, terkejut, jijik, dan netral. Variasi ekspresi wajah yang lebih kompleks, seperti rasa lelah atau kebingungan, tidak akan dianalisis dalam penelitian ini.

2. Kondisi Lingkungan Antar Kelas

Penelitian ini dilakukan dalam lingkungan kelas yang terkontrol dengan pencahayaan yang cukup. Pengaruh pencahayaan yang buruk, sudut pengambilan gambar yang ekstrem, atau gangguan visual seperti objek yang menghalangi wajah peserta didik tidak menjadi fokus dalam penelitian ini.

3. Waktu Pemrosesan

Penelitian ini berfokus pada evaluasi performa model berdasarkan mean Average Precision (mAP) dan waktu inferensi, diuji pada perangkat berbasis GPU (Google Colaboratory dengan spesifikasi Tesla T4) dan CPU (laptop peneliti) untuk mendukung aplikasi real-time. Proses pelatihan dan fine-tuning dilakukan menggunakan Google Colaboratory dengan GPU Tesla T4. Penelitian ini tidak mempertimbangkan efisiensi komputasi, seperti konsumsi memori atau daya, pada perangkat edge, sehingga hasil tidak dapat digeneralisasikan untuk skenario di luar lingkungan pengujian GPU dan CPU.

1.6. Sistematika Penulisan Skripsi

Sistem penulisan skripsi berfungsi sebagai panduan penulisan agar dalam penulisan ini lebih terarah. Oleh karena itu, skripsi ini dibagi menjadi beberapa bab. Struktur organisasi skripsi ini adalah sebagai berikut:

1. BAB I (Pendahuluan)

Bab ini menguraikan latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Pendahuluan berfungsi sebagai pengantar untuk menjelaskan pokok bahasan yang akan diteliti serta pentingnya penelitian ini dilakukan.

2. BAB II (Tinjauan Pustaka)

Bab ini berisi kajian teori yang mendukung penelitian, termasuk teori-teori mengenai pengenalan ekspresi wajah, transfer learning, dan teknologi Vision Transformer (ViT). Selain itu, bab ini juga membahas penelitian-penelitian terdahulu yang relevan dan bagaimana penelitian ini melengkapi atau memperluas hasil-hasil yang sudah ada.

3. BAB III (Metode Penelitian)

Bab ini menjelaskan metode yang digunakan dalam penelitian, mencakup pengumpulan data, pra-pemrosesan data, pengembangan model pengenalan ekspresi wajah berbasis Vision Transformer (ViT) dan metode YOLO, serta teknik pelatihan dan evaluasi. Bab ini juga memaparkan rancangan eksperimen, alat dan bahan yang digunakan, serta metode evaluasi performa model untuk mendukung pemantauan keadaan emosional peserta didik di lingkungan kelas.

4. BAB IV (Hasil dan Pembahasan)

Bab ini menyajikan hasil implementasi dan evaluasi model pengenalan ekspresi wajah di lingkungan kelas. Data hasil eksperimen dianalisis untuk menjawab rumusan masalah, dengan membandingkan performa pendekatan berbasis Vision Transformer (ViT) dan metode YOLO berdasarkan akurasi klasifikasi dan waktu inferensi. Pembahasan juga mencakup faktor-faktor yang memengaruhi performa model dalam konteks pendidikan.

5. BAB V (Kesimpulan dan Saran)

Bab ini menyimpulkan temuan utama dari penelitian dan menghubungkannya dengan tujuan yang telah ditetapkan sebelumnya. Selain itu, bab ini memberikan saran bagi penelitian lanjutan serta aplikasi praktis dari hasil penelitian ini di bidang pendidikan, khususnya terkait pengenalan ekspresi wajah di ruang kelas.