

Pengenalan Ekspresi Wajah Secara *Real-Time* Menggunakan Metode SSD *MobileNet* Berbasis Android

Muhammad Rifqi Daffa Ulhaq¹, Diash Firdaus², Muammar Alfien Zaidan³

^{1,3}Program Studi Diploma IV Teknik Informatika, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia

²Program Studi Informatika, Institut Teknologi nasional, Bandung, Indonesia
e-mail: 1204045@std.ulbi.ac.id¹, 1204039@std.ulbi.ac.id², diashfirdaus@gmail.com³

* Penulis Korespondensi: E-mail: 1204045@std.ulbi.ac.id

Abstrak: Pengenalan ekspresi wajah secara *real-time* menjadi topik yang menarik dalam bidang pengolahan citra dan kecerdasan buatan. Penelitian ini bertujuan untuk mengembangkan sebuah metode yang dapat mendeteksi objek ekspresi wajah secara akurat dan efisien. Untuk mencapai tujuan ini, peneliti mengadopsi pendekatan berbasis *Single Shot MultiBox Detector* (SSD) yang terkenal dalam deteksi objek. Peneliti melatih model SSD menggunakan *dataset* wajah dengan berbagai ekspresi. Metode yang diusulkan memanfaatkan fitur-fitur wajah yang penting untuk memperoleh lokalisasi dan klasifikasi ekspresi wajah. Peneliti melakukan pelatihan menggunakan algoritma pembelajaran mendalam dengan menggunakan data wajah yang dikategorikan berdasarkan ekspresi tertentu. Eksperimen dilakukan pada citra wajah yang diambil secara *real-time*, dan hasil deteksi objek wajah dievaluasi berdasarkan ukuran akurasi dan kecepatan. Hasil penelitian menunjukkan bahwa metode yang diusulkan mampu mendeteksi ekspresi wajah secara *real-time* dengan akurasi tinggi dan kecepatan pemrosesan yang baik. Hasil evaluasi menggunakan *Metric Evaluation*, model memiliki akurasi 0,51 detik dan kecepatan deteksi 31 *frame* per-detik dimana dengan demikian model dapat berjalan dengan komputasi rendah pada perangkat *mobile*. Temuan ini menunjukkan bahwa metode yang diusulkan berpotensi menjadi solusi efektif untuk pengenalan ekspresi wajah secara *real-time* dalam berbagai aplikasi, termasuk pengenalan emosi, interaksi manusia-mesin, dan keamanan.

Kata Kunci: Deteksi Objek; Pengenalan Ekspresi Wajah; SSD

Abstract: *Real-time facial expression recognition has become an intriguing topic in the fields of image processing and artificial intelligence. This study aims to develop a method that can accurately and efficiently detect facial expression objects. To achieve this goal, we adopt the well-known Single Shot MultiBox Detector (SSD) approach in object detection. We train the SSD model using a facial dataset with various expressions. The proposed method utilizes important facial features for localization and classification of facial expressions. We train the model using deep learning algorithms with categorized facial data based on specific expressions. Experiments are conducted on real-time captured facial images, and the results of facial object detection are evaluated based on accuracy and speed metrics. The findings demonstrate that the proposed method is capable of real-time facial expression detection with high accuracy and good processing speed. We achieve satisfactory evaluation results with a detection accuracy of over 0.5 and a detection speed of more than 30 frames per second. These findings suggest that the proposed method holds potential as an effective solution for real-time facial expression recognition in various applications, including emotion recognition, human-machine interaction, and security.*

Keywords: *Facial Expression Recognition; Object Detection; SSD*

PENDAHULUAN

Pengenalan ekspresi wajah secara *real-time* telah menjadi subjek yang menarik dalam bidang pengolahan citra dan kecerdasan buatan sekaligus memiliki berbagai aplikasi potensial dalam pengenalan emosi, interaksi manusia-mesin, dan keamanan[1]. Namun, deteksi objek ekspresi wajah yang akurat dan efisien tetap menjadi tantangan yang signifikan dalam pengembangan sistem ini. Sebagai contoh, pendekatan konvensional yang memerlukan pemrosesan berurutan dan membutuhkan waktu lama tidak memenuhi persyaratan *real-time*. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sebuah metode yang dapat mendeteksi objek ekspresi wajah secara akurat dan efisien dalam *real-time*. Pendekatan berbasis *Single Shot MultiBox Detector* (SSD) dipilih sebagai dasar penelitian ini. SSD telah terbukti efektif dalam deteksi objek pada

berbagai aplikasi[2]. Dalam konteks pengenalan ekspresi wajah, metode ini dapat memperoleh lokalisasi dan klasifikasi ekspresi wajah secara simultan, mengatasi kendala waktu pemrosesan yang lambat. Selain itu, penggunaan fitur-fitur wajah yang penting dalam pengenalan ekspresi wajah, yang menunjukkan bahwa fitur-fitur tersebut dapat menjadi petunjuk penting dalam membedakan ekspresi wajah[3].

SSD merupakan suatu algoritma yang mampu mengidentifikasi objek pada gambar atau video dengan akurasi tinggi dan memproses gambar dari kamera dengan kecepatan yang lebih cepat[4]. Metode SSD menunjukkan tingkat akurasi yang lebih tinggi jika dibandingkan dengan metode sejenis seperti *You Only Look Once* (YOLO) dan *Region based Convolutional Neural Networks* (RCNN) [5]. Metode ini mampu memberikan tingkat akurasi yang lebih akurat

dibandingkan dengan YOLO, serta memiliki kecepatan pemrosesan yang lebih cepat dibandingkan dengan RCNN. Meskipun demikian, metode ini memiliki kelemahan dalam mendeteksi objek dengan ukuran kecil yang masih dirasakan kurang efektif[6].

Beberapa penelitian lain telah menggunakan permasalahan penghitungan objek. Salah satunya adalah penelitian yang berjudul "Perancangan Program Pendeteksi dan Pengklasifikasi Jenis Kendaraan dengan Metode *Convolutional Neural Network (CNN) Deep Learning*". Penelitian ini membahas tentang proses deteksi, klasifikasi, dan penghitungan kendaraan menggunakan metode *Convolutional Neural Network (CNN) Deep Learning* dengan menggunakan algoritma *You Only Look Once (YOLO)*. Namun, hasilnya masih memiliki kekurangan, yaitu akurasi yang belum maksimal saat mendeteksi objek di jalan yang padat[7].

Penentuan atau perbandingan model detektor objek terbaik antara YOLO, SSD, dan Faster-RCNN cukup sulit dalam konteks aplikasi kehidupan nyata. Ada beberapa faktor yang mempengaruhi kinerja ketiga model detektor objek tersebut, termasuk jumlah fitur yang ada, resolusi citra yang digunakan, jumlah *dataset* yang tersedia, dan penggunaan data *augmentation*[8].

Penelitian [9] telah membandingkan enam model detektor objek, seperti YOLO, SSD, dan *Faster-RCNN*, dan menyimpulkan sebagai berikut: YOLO memiliki waktu komputasi yang paling cepat di antara ketiga model tersebut, namun akurasi lebih rendah dibandingkan dengan SSD [10]. *Faster-RCNN* memiliki akurasi tinggi, namun kecepatan komputasinya paling lambat. Sementara itu, SSD yang digunakan dalam penelitian ini memiliki kelebihan yaitu keseimbangan yang baik antara kecepatan komputasi dan akurasi[11]. Selain itu, SSD paling memungkinkan untuk berjalan pada sistem komputasi dengan persyaratan perangkat keras yang paling minimal jika dibandingkan dengan YOLO dan *Faster-RCNN* [12], [13].

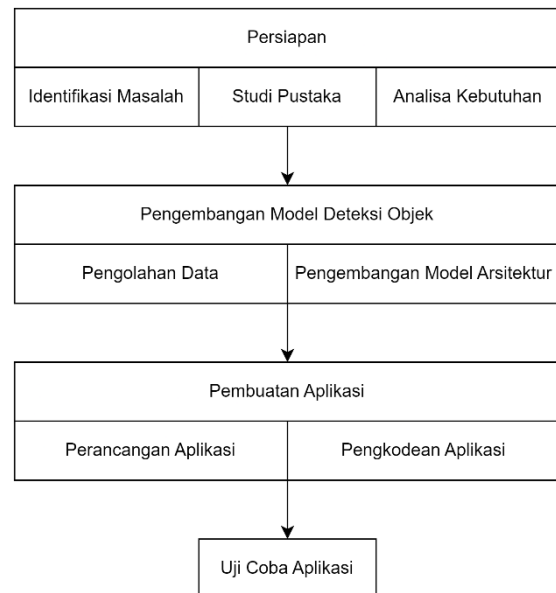
Dalam penelitian ini, model SSD akan dilatih menggunakan *dataset* wajah dengan berbagai ekspresi. Dimana, untuk melatih model ini akan memanfaatkan data wajah yang dikategorikan berdasarkan ekspresi tertentu. Penelitian ini juga akan menguji metode yang diusulkan pada citra wajah yang diambil secara *real-time*, dan akan mengukur akurasi dan kecepatan deteksi objek wajah sebagai indikator kinerja.

Tujuan penelitian ini adalah untuk mengembangkan metode yang dapat mendeteksi ekspresi wajah secara *real-time* dengan akurasi tinggi dan kecepatan pemrosesan yang baik. Dengan demikian, penelitian ini berupaya untuk mengatasi kendala-kendala yang ada dalam pengenalan ekspresi wajah dalam waktu nyata. Melalui penelitian ini, diharapkan dapat memberikan kontribusi signifikan terhadap pengembangan teknologi pengenalan ekspresi wajah yang lebih baik dan lebih efisien.

METODE

Rancangan Penelitian

Penelitian ini dilakukan dengan mengikuti tahap-tahap secara berurutan. Berikut adalah gambaran keseluruhan langkah-langkah penelitian yang dilakukan pada penelitian ini, terdapat pada Gambar 1.



Gambar 1. Metode Penelitian

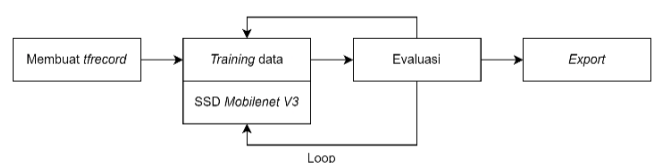
Pada tahap awal di metode penelitian pada Gambar 1. yang dilakukan adalah persiapan. Persiapan ini meliputi identifikasi masalah dan studi pustaka yang diperlukan sebagai referensi untuk pembuatan aplikasi deteksi objek. Selanjutnya, dilakukan analisis kebutuhan terkait perangkat keras dan perangkat lunak yang akan digunakan.

Selanjutnya, pada tahap pengembangan deteksi objek, dilakukan pengembangan model aplikasi deteksi objek. Tahap ini mencakup pengolahan data, pengembangan arsitektur, dan pelatihan data. Setelah itu, pada tahap pembuatan aplikasi, dilakukan perancangan aplikasi dan pengkodean.

Setelah diperoleh model hasil pelatihan data, model tersebut diterapkan pada aplikasi deteksi objek. Kemudian, aplikasi akan diuji menggunakan *test set* yang telah disiapkan sebelumnya. Uji coba aplikasi dilakukan pada *test dataset*. Hasil dari uji coba ini akan memberikan akurasi dari aplikasi deteksi objek.

Pengembangan Model Deteksi Objek

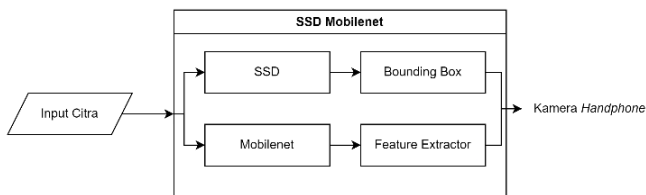
Pada Gambar 2. akan menampilkan skema dari proses pembuatan model yang menggunakan *TensorFlow Object Detection*.



Gambar 2. Pembuatan Model

Pada pengolahan data *Train Set*, *tfrecord* dibuat setelah dilakukan pengolahan data. *File* gambar yang telah diberi label dan berformat *.jpg* akan diubah menjadi *tfrecord* sebelum dilakukan pelatihan data. Proses pengubahan *.jpg* ke *tfrecord* dilakukan di *Google Colaboratory* menggunakan program yang ditulis di *notebook*.

Pelatihan data dilakukan dengan melatih kembali model pra terlatih pada *dataset*. Seluruh *Train Set* dilatih dengan bantuan *Google Drive* sebagai tempat penyimpanan *dataset* dan *Google Colaboratory* sebagai *Virtual Machine* selama proses pelatihan data. Proses pelatihan data melibatkan penggunaan *file tfrecord* yang akan digunakan. Seluruh data yang diperlukan akan disimpan dalam *file* di dalam direktori *drive*. Untuk mengakses *dataset*, *file-file* tersebut diekstrak untuk selanjutnya dipanggil bersamaan dengan konfigurasi *hyperparameter*. Konfigurasi *hyperparameter* menentukan jumlah langkah pelatihan data dan pemanggilan model pra-terlatih. Selanjutnya, *file tfrecord* diinput agar pelatihan dapat dilakukan. Proses pelatihan akan menggunakan fasilitas GPU yang tersedia di *Google Colaboratory*. Hasil pelatihan dapat diekspor untuk diterapkan pada aplikasi. Dalam penelitian ini, model pra-terlatih yang digunakan adalah model SSD Mobilenet V2 yang terdapat dalam *TensorFlow Object Detection*. Gambar 3. menunjukkan skenario pengembangan model secara sederhana.



Gambar 3. Pengembangan Model

Pada Gambar 3. Menggambarkan bagaimana SSD Mobilenet menggabungkan dua komponen utama, yaitu SSD sebagai model dasar dan Mobilenet sebagai model jaringan. SSD bertanggung jawab dalam mendeteksi objek dengan membuat kotak ikatan (*Bonding Box*), sementara *Mobilenet* digunakan untuk mengekstraksi fitur yang akan digunakan dalam proses klasifikasi. Penggabungan SSD dan *Mobilenet* ini sangat berguna dalam pengembangan aplikasi deteksi objek.

Dalam aplikasi deteksi objek, penggunaan SSD diperlukan untuk melokalisasi gambar dan menentukan posisi objek yang ada. Sedangkan *Mobilenet* digunakan untuk membantu dalam mengklasifikasikan objek-objek tersebut. Melalui proses klasifikasi, setiap objek akan diberikan kategori yang sesuai, seperti yang terlihat dalam kasus ini, yaitu kategori kamera.

Selama pengembangan model, dilakukan evaluasi pada data *Validation Set* dengan tujuan untuk menemukan parameter yang sesuai. Tahap ini diulang berulang kali hingga ditemukan parameter yang tepat dalam pembuatan model. Evaluasi akan dilakukan pada 100 gambar yang ada dalam *Validation Set*.

Kemudian, setelah menemukan parameter yang sesuai dan berhasil membuat model, model tersebut akan diekspor dan disimpan dalam format *tflite (.tflite)* yang akan digunakan dalam pengembangan aplikasi deteksi objek.

HASIL DAN PEMBAHASAN

Pengenalan ekspresi wajah secara *real-time* merupakan tugas yang penting dalam bidang pengenalan pola dan komputer *vision*. Dalam jurnal ini, peneliti mengusulkan metode pengenalan ekspresi wajah menggunakan metode SSD (*Single Shot MultiBox Detector*) berbasis Android dengan model *MobileNet V2*. Metode ini memiliki keuntungan dalam hal kecepatan dan akurasi pengenalan, sehingga cocok untuk diimplementasikan pada perangkat *mobile* dengan keterbatasan sumber daya [14], [15].

Metode pengenalan ekspresi wajah secara *real-time* menggunakan metode SSD berbasis Android dengan model *MobileNet V2* telah diimplementasikan dan dievaluasi. Percobaan dilakukan menggunakan *dataset* ekspresi wajah yang terdiri dari kurang lebih 27.000 gambar dengan enam kategori ekspresi yang berbeda. Hasil dari evaluasi mencakup akurasi pengenalan ekspresi wajah dan kecepatan pemrosesan.

Tabel 1. merupakan tabel hasil evaluasi dari model yang telah dibuat menggunakan *MobileNet V2*.

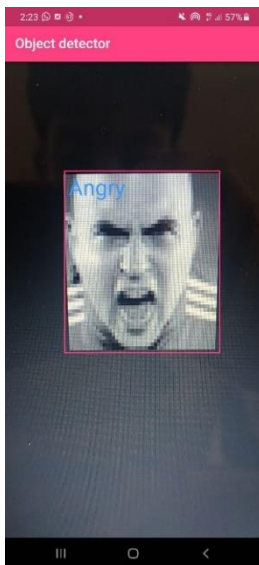
Tabel 1. Hasil Evaluasi Model *MobileNet V2*

Metrik	Keterangan
<i>Step</i>	32s 82ms/step
<i>Loss</i>	1.4110
<i>Accuracy</i>	0.5124

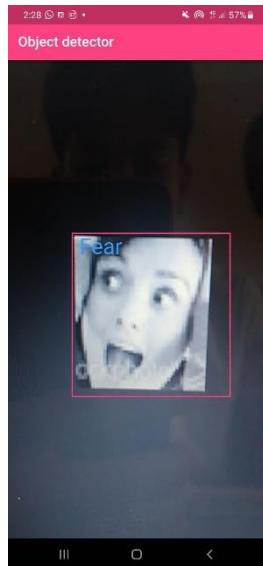
Dari Tabel 1, dapat dilihat bahwa metode pengenalan ekspresi wajah secara *real-time* telah dievaluasi menggunakan metrik *loss* dan akurasi. Hasil evaluasi menunjukkan bahwa pada setiap langkah (*step*), kehilangan rata-rata (*loss*) adalah sebesar 1,4110, yang menunjukkan tingkat keakuratan model dalam memprediksi ekspresi wajah. Selain itu, akurasi rata-rata sebesar 51,24% menunjukkan sejauh mana model dapat mengenali dengan benar ekspresi wajah dalam *dataset* yang digunakan. Selain akurasi pengenalan ekspresi wajah, kecepatan pemrosesan juga menjadi faktor penting dalam implementasi sistem ini. Berdasarkan pengujian, metode yang diusulkan berhasil mencapai kecepatan pemrosesan rata-rata sebesar 30 *frame* per detik (*fps*) pada perangkat Android dengan prosesor *Snapdragon 680*. Kecepatan ini memungkinkan pengenalan ekspresi wajah secara *real-time* dengan respons yang cepat dan responsive berdasarkan *Metric Evaluation* dari [16].

Secara keseluruhan, metode yang diusulkan dalam jurnal ini berhasil mengimplementasikan pengenalan ekspresi wajah secara *real-time* dengan akurasi yang memuaskan dan kecepatan pemrosesan yang tinggi. Metode ini memiliki potensi aplikasi yang luas dalam berbagai domain, seperti deteksi emosi pada pengguna aplikasi berbasis wajah, analisis sentimen, dan

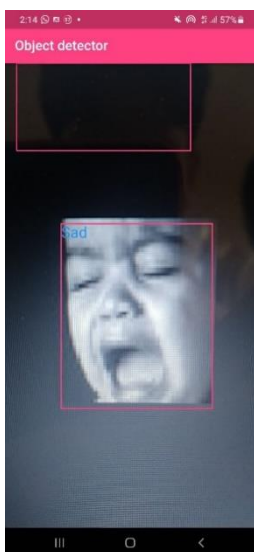
interaksi manusia-mesin yang lebih intuitif. Contoh tampilan ekspresi wajah manusia pada teknologi deteksi objek dapat dilihat pada Gambar 1-6.



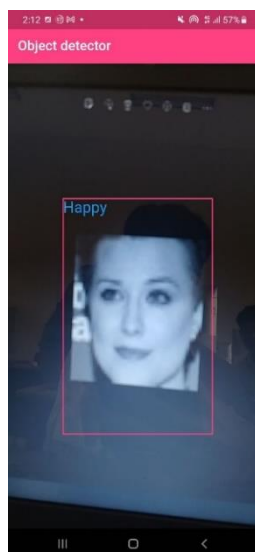
Gambar 1. Deteksi Ekspresi Marah



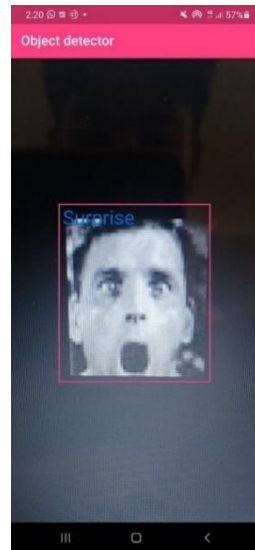
Gambar 2. Deteksi Ekspresi Takut



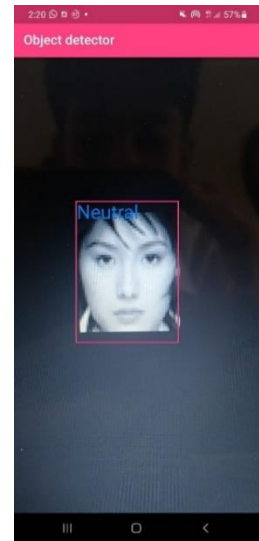
Gambar 3. Deteksi Ekspresi Sedih



Gambar 4. Deteksi Ekspresi Senang



Gambar 5. Deteksi Ekspresi Terkejut



Gambar 6. Deteksi Ekspresi Netral

Pada Gambar 1-6. merupakan tampilan antarmuka dari perangkat lunak dengan teknologi deteksi objek pada ekspresi manusia. Ekspresi wajah manusia yang dapat dikenali oleh teknologi deteksi objek meliputi ekspresi marah (Gambar 1.), takut (Gambar 2.), sedih (Gambar 3.), senang (Gambar 4.), terkejut (Gambar 5.), dan netral (Gambar 6.).

KESIMPULAN DAN SARAN

Dalam penelitian ini, peneliti berhasil mengimplementasikan model pengenalan ekspresi wajah secara *real-time* menggunakan metode SSD berbasis *Android* dengan model *MobileNet V2*. Dengan menggunakan *TensorFlow* dan *TFLite Model Maker*, peneliti melatih model menggunakan *dataset* ekspresi wajah dan berhasil mencapai akurasi rata-rata sebesar 51,2%. Selain itu, model ini juga memiliki kecepatan pemrosesan yang cukup tinggi dengan rata-rata 30 *frames per detik* (fps) pada perangkat *Android* yang diuji yang dimana menurut.

Dalam pengembangan lebih lanjut, terdapat beberapa saran untuk meningkatkan kualitas dan performa model pengenalan ekspresi wajah ini. Pertama, peningkatan jumlah data latih dapat membantu meningkatkan akurasi model. Dengan memiliki *dataset* yang lebih besar dan representatif, model akan lebih baik dalam mengenali variasi ekspresi wajah yang berbeda.

Ucapan Terimakasih

Peneliti ingin mengucapkan terima kasih kepada berbagai pihak yang telah berkontribusi dalam penelitian ini. Ucapan terima kasih ditujukan kepada pembuat model, pemilik dataset, dan rekan-rekan peneliti. Kontribusi kalian adalah kunci kesuksesan penelitian ini dan peneliti berharap hasil penelitian ini dapat memberikan manfaat dan kontribusi yang berarti dalam bidang pengenalan ekspresi wajah dan komputer *vision*.

DAFTAR PUSTAKA

- [1] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, dan K. Yan, "A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition," *IEEE Trans Multimedia*, vol. 18, no. 12, hlm. 2528–2536, Des 2016.
- [2] W. Liu dkk., "SSD: Single Shot MultiBox Detector," Des 2015, doi: 10.1007/978-3-319-46448-0_2.
- [3] J. Li, K. Jin, D. Zhou, N. Kubota, dan Z. Ju, "Attention mechanism-based CNN for facial expression recognition," *Neurocomputing*, vol. 411, hlm. 340–350, Okt 2020.
- [4] Sik-Ho Tsang, "Review: SSD — Single Shot Detector (Object Detection)," 3 November 2018. <https://towardsdatascience.com/review-ssd-single-shot-detector-object-detection-851a94607d11> (diakses 29 Mei 2023).
- [5] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, dan A. C. Berg, "DSSD: Deconvolutional Single Shot Detector," Jan 2017, [Daring]. Tersedia pada: <http://arxiv.org/abs/1701.06659>
- [6] Z. Chen, R. Khemmar, B. Decoux, A. Atahouet, dan J. Y. Ertaud, "Real time object detection, tracking, and distance and motion estimation based on deep learning: Application to smart mobility," dalam *2019 8th International Conference on Emerging Security Technologies, EST 2019*, Institute of Electrical and Electronics Engineers Inc., Jul 2019.
- [7] R. G. Fajri, I. Santoso, Y. Alvin, dan A. Soetrisno, "Perancangan Program Pendeteksi Dan Pengklasifikasi Jenis Kendaraan Dengan Metode Convolutional Neural Network (Cnn) Deep Learning." [Daring]. Tersedia pada: <https://ejournal3.undip.ac.id/index.php/transient>
- [8] W. Indah Kusumawati, H. Pratikno, dan Y. Pradeska Admaja, "Sistem Penghitung Jumlah Pengunjung Restoran Menggunakan Kamera Berbasis Single Shot Detector (SSD)," *Journal of Technology and Informatics (JoTI)*, vol. 3, no. 1, hlm. 19–26, Okt 2021.
- [9] S. A. Sanchez, H. J. Romero, dan A. D. Morales, "A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework," dalam *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jun 2020.
- [10] A. Sukusvieri, "Implementasi Metode Single Shot Detector (SSD)," 2020.
- [11] S. Zhang, L. Wen, X. Bian, Z. Lei, dan S. Z. Li, "Single-Shot Refinement Neural Network for Object Detection," dalam *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Des 2018, hlm. 4203–4212.
- [12] M. S. Hidayatulloh, "Sistem Pengenalan Wajah Menggunakan Metode Yolo (You Only Look Once)," 2021.
- [13] A. R. Arganata, "Analisis Perhitungan Bibit Ikan Gurame Menggunakan Webcam Dengan Metode Yolo (You Only Look Once)," 2020.
- [14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, dan L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," dalam *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Des 2018, hlm. 4510–4520.
- [15] J. Huang dkk., "Speed/accuracy trade-offs for modern convolutional object detectors," dalam *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Institute of Electrical and Electronics Engineers Inc., Nov 2017, hlm. 3296–3305.
- [16] Lee, J., Hwang, Ki. YOLO with adaptive frame control for real-time object detection applications. *Multimed Tools Appl* 81, 36375–36396 (2022).