

BAB III

METODE PENELITIAN

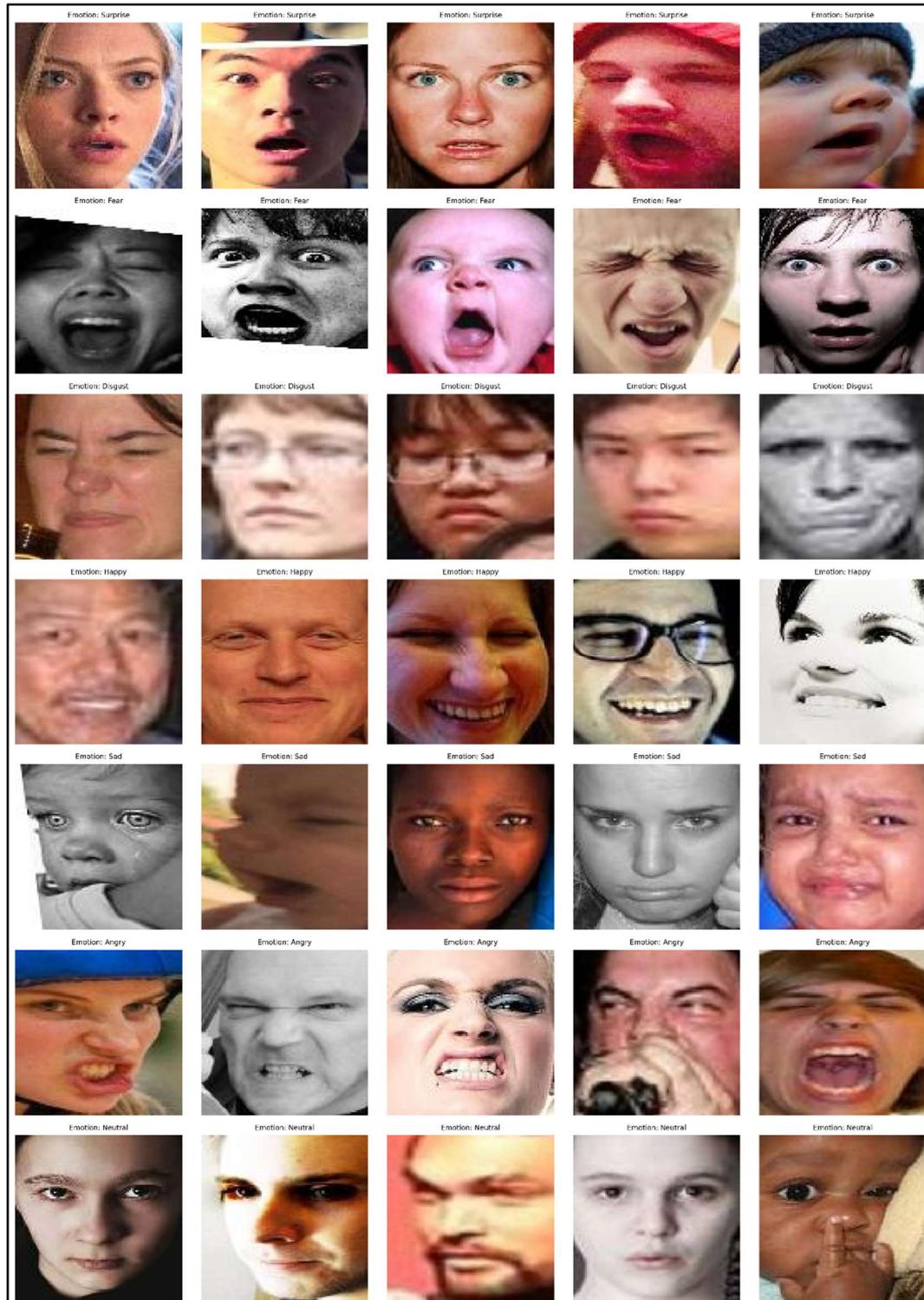
Bab ini menguraikan secara sistematis rancangan penelitian yang dilakukan, mulai dari pendekatan dan metode yang digunakan, teknik pengumpulan data, hingga tahapan pra-pemrosesan data yang dilakukan untuk mempersiapkan data sebelum digunakan dalam pelatihan model. Bab ini juga mencakup pembahasan mengenai arsitektur model yang diimplementasikan, serta alur penelitian secara keseluruhan mulai dari studi literatur hingga tahap penulisan skripsi, lengkap dengan parameter pelatihan dan metrik evaluasi yang digunakan untuk menilai performa model.

3.1. Rancangan Penelitian

Penelitian ini dilakukan dengan menggunakan pendekatan eksperimen kuantitatif. Pendekatan ini bertujuan untuk menguji performa model Hybrid Vision Transformer (HybridViT) dalam mengklasifikasikan ekspresi wajah peserta didik berdasarkan data citra di ruang kelas dan membuat model pembanding YOLOv11 (You Only Look Once). Proses penelitian terdiri atas beberapa tahapan utama, yaitu pengumpulan dataset, pra-pemrosesan data, pelatihan model, pengujian model, dan evaluasi hasil.

3.2. Pengumpulan Data

Penelitian ini memanfaatkan tiga dataset utama yang masing-masing berperan dalam proses pelatihan, fine-tuning, dan evaluasi sistem pengenalan ekspresi wajah berbasis deep learning. Pemilihan dataset dilakukan secara selektif berdasarkan kesesuaian karakteristik data terhadap kebutuhan model, kualitas anotasi, serta kemudahan aksesibilitas.



Gambar 3.1. Cuplikan Isi Dataset RAF-DB

Dataset pertama yang digunakan adalah Real-world Affective Face Database (RAF-DB) yang dipublikasikan oleh Alok (2023) melalui platform Kaggle. Dataset ini memuat kurang lebih 15.300 citra wajah yang dikategorikan ke

dalam tujuh emosi dasar, yaitu bahagia, sedih, marah, takut, terkejut, jijik, dan netral. RAF-DB dipilih karena menyajikan variasi ekspresi wajah yang subtil dan realistis, sehingga sangat representatif dengan kondisi emosi peserta didik di dalam kelas, yang cenderung tidak terlalu ekspresif. Selain itu, gambar dalam dataset ini memiliki resolusi relatif tinggi (100×100 piksel) dan berwarna, yang meningkatkan kualitas input visual untuk model Hybrid Vision Transformer (HybridViT). Aksesibilitasnya melalui Kaggle juga memudahkan integrasi ke dalam pipeline pelatihan berbasis cloud seperti Google Colaboratory.



Gambar 3.2. Cuplikan Isi Dataset Face Detection Dataset

Dataset kedua adalah Face Detection Dataset yang dikembangkan oleh Elmenshawii (2023) dan juga tersedia di platform Kaggle. Dataset ini terdiri dari sekitar 16.700 gambar yang memuat satu atau lebih wajah dengan anotasi berbasis format YOLO (You Only Look Once), sehingga cocok untuk pelatihan ulang (fine-tuning) model deteksi wajah YOLOv11s. Dataset ini memiliki keunggulan dalam konsistensi anotasi dan kualitas gambar yang baik, dengan bounding box yang akurat, menjadikannya kandidat tepat untuk pengembangan sistem deteksi wajah real-time dalam lingkungan kelas.



Gambar 3.3. Cuplikan isi dataset Facial Expression in Classroom

Dataset ketiga, Facial Expression in Classroom Dataset, dikembangkan dan dipublikasikan oleh Khaairi (2025) melalui platform Roboflow. Dataset ini terdiri dari tiga sesi pengambilan data, yaitu Sesi 1, Sesi 2, dan Sesi 3, dengan total 1.162 gambar setelah penghapusan manual pada Sesi 3. Setiap gambar dilengkapi dengan

bounding box dan label emosi yang mencerminkan tujuh ekspresi wajah dasar (Happy, Sad, Angry, Fear, Surprise, Disgust, Neutral). Ekspresi dalam dataset ini bersifat natural dan subtil, mencerminkan kondisi emosional peserta didik dalam situasi belajar di ruang kelas. Sesi 1 dan Sesi 2 digunakan untuk fine-tuning model dual-stage (HybridViT untuk pengenalan ekspresi wajah + YOLOv11s untuk deteksi wajah) dan model single-stage (YOLOv11s untuk deteksi emosi), guna meningkatkan akurasi dan generalisasi model dalam konteks kelas. Sesi 3, yang memiliki jumlah wajah lebih banyak dibandingkan Sesi 1 dan 2, digunakan secara eksklusif untuk pengujian performa akhir model berdasarkan mean Average Precision (mAP) dan waktu inferensi. Selain itu, terdapat set pengujian tambahan di Roboflow, bernama Test, yang hanya berisi Sesi 3 dengan jumlah 536 gambar, digunakan untuk evaluasi performa model pada skenario pengujian terpisah. Dataset ini dipilih untuk tahap fine-tuning dan pengujian karena relevansinya dengan tujuan akhir sistem. Karakteristik emosi serta konteks pengambilan datanya sejalan dengan penerapan utama model, yaitu Pengenalan Ekspresi Wajah (FER) di dalam ruang kelas.

Ketiga dataset tersebut dipilih untuk memenuhi kebutuhan evaluasi sistem secara holistik—mulai dari pelatihan awal pada data ekspresi umum, pelatihan deteksi wajah multivarian, hingga pengujian pada konteks nyata ruang kelas. Pendekatan ini diharapkan mampu memastikan robustitas dan relevansi model terhadap lingkungan aplikasi yang dituju.

3.3. Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk mempersiapkan dataset agar sesuai dengan kebutuhan model pembelajaran mesin yang digunakan dalam penelitian ini. Pra-pemrosesan mencakup transformasi gambar, augmentasi data, normalisasi, dan pengorganisasian data ke dalam struktur yang dapat diproses oleh model. Berikut adalah rincian tahap pra-pemrosesan untuk masing-masing program yang dikembangkan:

1. Pelatihan Model Hybrid Vision Transformer (HybridViT) ResNet-50 (Pengenalan Ekspresi Wajah)

Pra-pemrosesan data pada program ini dirancang untuk mempersiapkan dataset RAF-DB untuk pelatihan model Hybrid Vision Transformer (HybridViT) dengan backbone ResNet-50 untuk tugas pengenalan ekspresi wajah (Facial Expression Recognition, FER). Tahap pra-pemrosesan meliputi transformasi data sebagai berikut:

- 1) Transformasi Data Latih:
 - a) Resize Gambar: Gambar diubah ukurannya menjadi 224×224 piksel untuk memenuhi persyaratan input model.
 - b) Augmentasi Data: Dilakukan augmentasi untuk meningkatkan variasi data dan mencegah overfitting, meliputi:
 - i) RandomHorizontalFlip: Membalik gambar secara horizontal dengan peluang 50%.
 - ii) RandomRotation: Rotasi acak hingga 10 derajat untuk memperkaya variasi orientasi gambar.
 - c) Normalisasi: Nilai piksel dinormalisasi menggunakan rata-rata [0.485, 0.456, 0.406] dan deviasi standar [0.229, 0.224, 0.225], sesuai dengan parameter pre-trained model ResNet-50.
 - d) Konversi ke Tensor: Gambar dikonversi ke format tensor PyTorch menggunakan fungsi ToTensor.
- 2) Transformasi Data Uji:

Transformasi hanya mencakup resize ke ukuran 224×224 piksel, konversi ke Tensor, dan normalisasi dengan parameter yang sama seperti data latih, tanpa augmentasi untuk menjaga integritas evaluasi.

2. Fine-tuning Model YOLOv11s (Deteksi Wajah)

Pra-pemrosesan data pada program ini bertujuan untuk mempersiapkan dataset Face Detection Dataset dari Kaggle untuk fine-tuning model YOLOv11s dalam tugas deteksi wajah. Pra-pemrosesan sebagian besar dikelola secara otomatis oleh pustaka Ultralytics YOLOv11, dengan detail sebagai berikut:

- 1) Transformasi Gambar Pelatihan:

- a) YOLOv11 secara otomatis melakukan *resize* gambar ke ukuran input 640×640 piksel dan normalisasi internal selama proses pelatihan.
 - b) Augmentasi data, seperti rotasi, *flipping*, dan *scaling*, diterapkan secara bawaan oleh YOLOv11 untuk meningkatkan robustitas model, tanpa transformasi eksplisit yang didefinisikan dalam kode.
- 2) Konfigurasi Dataset:
- Dataset diorganisasi menggunakan file konfigurasi dataset.yaml, yang mendefinisikan jalur data latih dan validasi serta kelas (face). Label *bounding box* dalam format YOLO (class_id, center_x, center_y, width, height) diverifikasi untuk memastikan konsistensi.
- 3) Pra-pemrosesan untuk Validasi:
- Untuk keperluan validasi, gambar dimuat dan diubah ke format RGB. Label *bounding box* dikonversi dari format YOLO ke koordinat piksel absolut (x_min, y_min, x_max, y_max) berdasarkan dimensi gambar untuk visualisasi dan evaluasi.

Tidak ada transformasi eksplisit tambahan yang didefinisikan, karena YOLOv11 mengelola pra-pemrosesan secara internal untuk memastikan efisiensi pelatihan.

3. Fine-tuning Pipeline Model Hybrid Vision Transformer (HybridViT) ResNet-50 (Pengenalan Ekspresi Wajah) dan YOLOv11s (Deteksi Wajah)

Pra-pemrosesan data pada program ini dirancang untuk mempersiapkan dataset Facial Expression in Classroom dari Roboflow untuk fine-tuning model HybridViT serta evaluasi pipeline deteksi wajah dan FER dalam lingkungan kelas. Tahap pra-pemrosesan mencakup:

- 1) Transformasi Data Fine-Tuning (Sesi 1 dan Sesi 2):
 - a) Data Latih:
 - i) Augmentasi Data: Menggunakan pustaka Albumentations untuk augmentasi yang lebih agresif, meliputi:
 - (1) A.Rotate: Rotasi acak hingga 15 derajat (peluang 50%).

- (2) A.RandomBrightnessContrast: Penyesuaian kecerahan dan kontras hingga 0.4 (peluang 50%).
 - (3) A.HueSaturationValue: Penyesuaian hue, saturasi, dan nilai hingga 20-30 (peluang 30%).
 - (4) A.RandomCrop: Pemotongan acak ke ukuran 224×224 piksel (peluang 30%).
 - (5) A.GaussNoise: Penambahan noise Gaussian dengan varians 10-50 (peluang 30%).
 - (6) A.Affine: Transformasi affine dengan translasi 10% dan scaling 0.9-1.1 (peluang 30%).
- ii) Normalisasi: Nilai piksel dinormalisasi dengan rata-rata [0.485, 0.456, 0.406] dan deviasi standar [0.229, 0.224, 0.225].
 - iii) Konversi ke Tensor: Gambar dikonversi ke tensor PyTorch menggunakan ToTensorV2.
- b) Data Validasi:
 - Hanya dilakukan normalisasi dengan parameter yang sama dan konversi ke Tensor, tanpa augmentasi untuk menjaga integritas evaluasi.
- 2) Pemotongan Wajah:
- a) Bounding box dari label YOLO digunakan untuk memotong area wajah dengan tambahan padding 20% guna menangkap konteks wajah secara optimal.
 - b) Jika ukuran wajah kurang dari 64×64 piksel, koordinat disesuaikan untuk memastikan ukuran minimal 64×64 piksel.
 - c) Area wajah yang dipotong diubah ke format BGR, kemudian diterapkan transformasi preprocess (resize ke 224×224 piksel, normalisasi, dan konversi ke tensor) untuk input ke model FER.
- 3) Pemuatan Data:
- a) Dataset dibagi menjadi data latih (80%) dan validasi (20%) menggunakan fungsi random_split untuk dataset Sesi 1 dan Sesi 2.
 - b) Fungsi custom_collate_fn diterapkan untuk mengelola struktur data kompleks yang mencakup gambar, label, bounding box, dan jalur file.

4. Fine-tuning Model YOLOv11s (Deteksi Emosi)

Pra-pemrosesan data pada program ini bertujuan untuk mempersiapkan dataset Facial Expression in Classroom Dataset dari Roboflow untuk fine-tuning model YOLOv11s dalam tugas deteksi emosi secara langsung (single-stage). Pra-pemrosesan sebagian besar dikelola secara otomatis oleh pustaka Ultralytics YOLOv11, dengan detail sebagai berikut:

- 1) Transformasi Gambar Pelatihan (Sesi 1 dan Sesi 2):
 - a) YOLOv11 secara otomatis melakukan resize gambar ke ukuran input 640×640 piksel dan normalisasi internal selama proses pelatihan.
 - b) Augmentasi data, seperti rotasi acak, flipping horizontal, dan scaling, diterapkan secara bawaan oleh YOLOv11 untuk meningkatkan robustitas model terhadap variasi ekspresi wajah dalam lingkungan kelas, tanpa transformasi eksplisit yang didefinisikan dalam kode.
- 2) Konfigurasi Dataset:
 - a) Dataset diorganisasi menggunakan file konfigurasi data.yaml, yang mendefinisikan jalur absolut ke direktori data latih (train/images) dan validasi (valid/images), jumlah kelas (nc=7), dan nama kelas (Happy, Sad, Angry, Fear, Surprise, Disgust, Neutral).
 - b) Dataset dibagi menjadi data latih (80%) dan validasi (20%) menggunakan pemisahan acak dengan random_state 42 untuk memastikan reproduktibilitas.

Tidak ada transformasi eksplisit tambahan yang didefinisikan, karena YOLOv11 mengelola pra-pemrosesan secara internal untuk memastikan efisiensi pelatihan dan evaluasi.

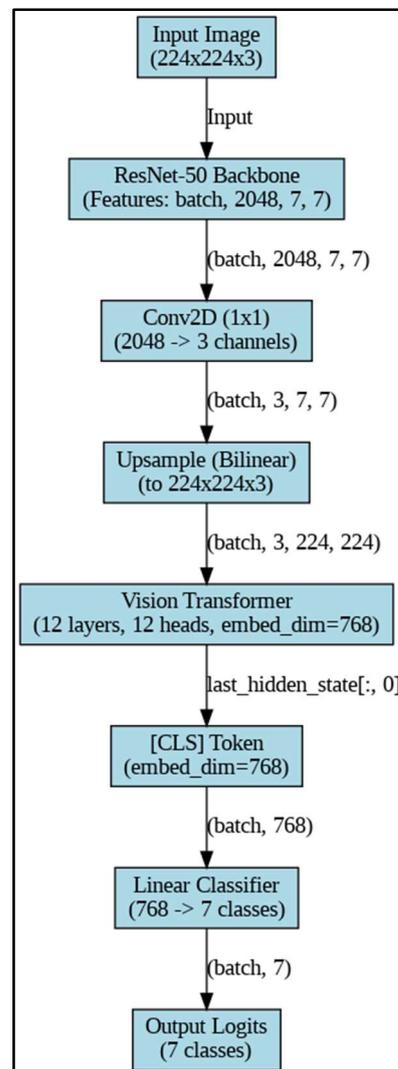
3.4. Arsitektur Model

Subbab ini menjelaskan arsitektur model yang digunakan dalam keempat program yang dikembangkan untuk tugas deteksi wajah dan pengenalan ekspresi wajah, dan deteksi emosi (single-stage) serta alur kerja masing-masing model. Arsitektur model dirancang untuk mengoptimalkan performa pada dataset spesifik dengan memanfaatkan pendekatan pembelajaran mendalam berbasis Convolutional

Neural Network (CNN) dan Vision Transformer (ViT). Berikut adalah rincian arsitektur dan alur kerja untuk masing-masing program.

1. Model Hybrid Vision Transformer (HybridViT) ResNet-50 (Pengenalan Ekspresi Wajah)

Model yang digunakan adalah HybridViT, sebuah arsitektur hibrida yang mengintegrasikan backbone ResNet-50 dengan *Vision Transformer* (ViT) untuk tugas klasifikasi tujuh kelas ekspresi wajah (*Surprise, Fear, Disgust, Happy, Sad, Angry, Neutral*). Struktur model terdiri dari komponen berikut:



Gambar 3.4. Diagram Arsitektur Model HybridViT ResNet-50

- 1) Backbone ResNet-50: Menggunakan ResNet-50 yang telah dilatih sebelumnya (*pretrained*) pada *dataset* ImageNet. Lapisan *fully connected* (FC) dan *average pooling* dihapus, sehingga menghasilkan fitur dengan dimensi (*batch*, 2048, 7, 7).
- 2) Adaptasi Saluran: Lapisan konvolusi 1×1 (`nn.Conv2d`) mengubah jumlah saluran dari 2048 menjadi 3 untuk menyesuaikan dengan *input* ViT (RGB).
- 3) Upsampling: Lapisan *upsampling* bilinear (`nn.Upsample`) mengubah dimensi spasial fitur menjadi 224×224 piksel, sesuai dengan ukuran *input* ViT.
- 4) Vision Transformer (ViT): Menggunakan konfigurasi ViT dengan parameter: `hidden_size=768`, `num_hidden_layers=12`, `num_attention_heads=12`, `intermediate_size=3072`, `image_size=224`, dan `patch_size=16`. ViT memproses fitur yang telah diadaptasi untuk menghasilkan representasi global melalui token [CLS].
- 5) Klasifikasi: Lapisan linear (`nn.Linear`) memetakan representasi token [CLS] (dimensi 768) ke tujuh kelas ekspresi wajah.

2. Model YOLOv11s (Deteksi Wajah)

Model yang digunakan adalah YOLOv11s, varian kecil (*small*) dari arsitektur YOLOv11 (You Only Look Once) yang dioptimalkan untuk deteksi wajah. YOLOv11s adalah model berbasis CNN yang dirancang untuk deteksi objek secara real-time dengan keseimbangan antara akurasi dan efisiensi komputasi. Arsitektur utama meliputi:

- 1) Backbone: Menggunakan struktur Cross Stage Partial Network (CSPNet) yang dimodifikasi, dengan penerapan blok C3k2 serta modul Spatial Pyramid Pooling – Fast (SPPF). Model ini juga mengintegrasikan mekanisme Cross-Stage Partial with Spatial Attention (C2PSA) untuk meningkatkan akurasi dalam mendeteksi fitur spasial penting pada objek wajah.
- 2) Neck: Menggunakan struktur Feature Pyramid Network (FPN) dan Path Aggregation Network (PAN) untuk mengintegrasikan fitur dari berbagai skala, memungkinkan deteksi wajah pada ukuran yang bervariasi.

- 3) Head: Lapisan deteksi yang menghasilkan prediksi bounding box, probabilitas kelas (face), dan skor kepercayaan (confidence score). Model diatur untuk mendeteksi satu kelas (face), dioptimalkan melalui fine-tuning pada dataset Face Detection Dataset.

3. Model YOLOv11s (Deteksi Emosi)

Model yang digunakan adalah YOLOv11s, varian kecil (small) dari arsitektur YOLOv11 (You Only Look Once) yang dioptimalkan untuk deteksi emosi secara langsung dalam pendekatan single-stage. YOLOv11s adalah model berbasis CNN yang dirancang untuk deteksi objek secara real-time dengan keseimbangan antara akurasi dan efisiensi komputasi. Arsitektur utama meliputi:

- 1) Backbone: Menggunakan struktur Cross Stage Partial Network (CSPNet) yang dimodifikasi, dengan penerapan blok C3k2 serta modul Spatial Pyramid Pooling – Fast (SPPF). Model ini juga mengintegrasikan mekanisme Cross-Stage Partial with Spatial Attention (C2PSA). Penggunaan modul-modul ini memungkinkan ekstraksi fitur yang lebih kaya terhadap ekspresi wajah.
- 2) Neck: Menggunakan struktur Feature Pyramid Network (FPN) dan Path Aggregation Network (PAN) untuk mengintegrasikan fitur dari berbagai skala, memungkinkan deteksi emosi pada ukuran yang bervariasi.
- 3) Head: Lapisan deteksi yang menghasilkan prediksi bounding box, probabilitas kelas untuk tujuh ekspresi wajah (Happy, Sad, Angry, Fear, Surprise, Disgust, Neutral), dan skor kepercayaan (confidence score). Model diatur untuk mendeteksi tujuh kelas emosi, dioptimalkan melalui fine-tuning pada dataset Facial Expression in Classroom.

3.5. Alur Penelitian

Penelitian ini mengikuti alur kerja yang sistematis untuk mengembangkan dan mengevaluasi sistem deteksi wajah dan pengenalan ekspresi wajah (Facial Expression Recognition (FER)) (dual-stage), dan deteksi emosi (single-stage) menggunakan pendekatan pembelajaran mendalam. Alur penelitian terdiri dari lima

tahap utama: studi literatur, pra-eksekusi, pelatihan, evaluasi, dan penulisan skripsi. Berikut adalah penjelasan rinci untuk masing-masing tahap, yang telah disesuaikan dengan implementasi dalam ketiga program penelitian.

1. Studi Literatur

Tahap ini bertujuan untuk membangun landasan teoretis dan metodologis penelitian. Langkah-langkah yang dilakukan meliputi:

- 1) Pengumpulan Referensi Ilmiah: Mengumpulkan jurnal, artikel, dan publikasi ilmiah yang relevan dengan topik deteksi wajah dan FER, dengan fokus pada pendekatan berbasis Convolutional Neural Network (CNN) dan Vision Transformer (ViT).
- 2) Analisis Penelitian Terkait: Mengkaji teori, konsep, dan teknologi yang mendasari FER, termasuk arsitektur model seperti ResNet, YOLO, dan ViT, serta evaluasi performa. Penelitian terkait dianalisis untuk memahami kelebihan, kekurangan, dan celah penelitian.
- 3) Perbandingan Penelitian: Membuat rangkuman yang mencakup hasil, kelebihan, dan keterbatasan penelitian sebelumnya, khususnya dalam hal akurasi, efisiensi, dan aplikabilitas pada skenario dunia nyata seperti lingkungan kelas. Perbandingan ini digunakan untuk merumuskan pendekatan penelitian yang optimal.

2. Pra-eksekusi

Tahap pra-eksekusi mencakup persiapan data, model, dan lingkungan komputasi untuk mendukung pelatihan dan evaluasi. Langkah-langkahnya adalah:

- 1) Pengumpulan Dataset:
 - a) Dataset RAF-DB diunduh dari Kaggle untuk pelatihan model FER (Hybrid Vision Transformer (HybridViT) ResNet-50).
 - b) Dataset Face-Detection-Dataset diunduh dari Kaggle untuk fine-tuning model deteksi wajah (YOLOv11s).
 - c) Dataset Facial-Expression-in-Classroom (Set utama (Sesi 1 dan 2) dan set testing (Sesi 3) diunduh dari Roboflow untuk fine-tuning dan

evaluasi pipeline FER (YOLOV11s dan HybridViT) dan model deteksi emosi (single-stage FER) (YOLOv11s) dalam lingkungan kelas

- d) Kode disiapkan untuk mengakses dataset melalui API Kaggle dan Roboflow, dengan verifikasi struktur data pada lingkungan Google Colaboratory. Dataset yang memerlukan penyesuaian terlebih dahulu disimpan dalam format .zip, kemudian diunggah ke lingkungan runtime untuk digunakan dalam proses pelatihan dan pengujian.
- 2) Perancangan Arsitektur Model:
- a) Model HybridViT dirancang dengan mengintegrasikan backbone ResNet-50 dan ViT untuk FER, digunakan pada program pertama dan ketiga.
 - b) Model YOLOv11s dipilih untuk deteksi wajah, digunakan pada program kedua dan sebagai bagian pipeline pada program ketiga, juga untuk deteksi emosi, digunakan pada program keempat.
 - c) Penentuan Pra-pemrosesan: Proses pra-pemrosesan dirancang, termasuk resize gambar, augmentasi data (RandomHorizontalFlip, RandomRotation, RandomBrightnessContrast, dll.), normalisasi, dan konversi ke Tensor, seperti dijelaskan pada Subbab 3.3.
 - d) Persiapan Kode Program: Kode program dikembangkan dalam format Jupyter Notebook untuk keperluan pelatihan, fine-tuning, dan evaluasi model. Proses eksekusi dilakukan pada platform Google Colaboratory dengan runtime GPU Tesla T4 untuk evaluasi berbasis GPU, serta menggunakan runtime Miniconda di laptop pribadi untuk evaluasi berbasis CPU. Berbagai dependensi yang diperlukan, seperti torch, transformers, ultralytics, dan albumentations, diinstal guna mendukung implementasi dan pengujian model.

3. Pelatihan

Tahap pelatihan melibatkan eksekusi program untuk melatih dan mengoptimalkan model. Langkah-langkahnya meliputi:

- 1) Persiapan Program: Kode diatur untuk memuat dataset, model, dan konfigurasi pelatihan (optimizer, loss function, scheduler). Dataset dimuat ke DataLoader dengan batch size 32.
- 2) Konfigurasi Runtime: Program dihubungkan dengan runtime GPU T4 pada Google Colaboratory untuk mempercepat pelatihan.
- 3) Eksekusi Pelatihan:
 - a) Program Pertama (Pelatihan Model FER): Model HybridViT dilatih pada dataset RAF-DB selama 20 epoch dengan CrossEntropyLoss, optimizer Adam (learning rate=1e-4), dan scheduler StepLR (step_size=10, gamma=0.1).
 - b) Program Kedua (Fine-Tuning Model Face Detection): Model YOLOv11s di-fine-tune pada dataset Face Detection Dataset selama 16 epoch dengan batch size 32, menggunakan konfigurasi bawaan Ultralytics dan early stopping (patience=2).
 - c) Program Ketiga (Fine-Tuning Model FER pada Dataset Classroom): Model HybridViT di-fine-tune pada dataset Facial Expression in Classroom selama maksimum 50 epoch dengan batch size 32, CrossEntropyLoss berbobot, optimizer Adam (learning rate=5e-5, weight_decay=5e-5), scheduler CosineAnnealingLR, dan early stopping (patience=5) berdasarkan mean Average Precision (mAP)
 - d) Program Keempat (Fine-Tuning Model Emotion Detection): Model YOLOv11s di-fine-tune pada dataset Facial-Expression-in-Classroom (Sesi 1 dan Sesi 2) selama 50 epoch dengan batch size 32, learning rate sebesar 0.001, dan menggunakan optimizer AdamW. Pelatihan menggunakan konfigurasi bawaan Ultralytics serta menerapkan early stopping dengan patience bernilai 5.
- 4) Pengaturan Hiperparameter: Hiperparameter seperti jumlah epoch, learning rate, dan batch size disesuaikan untuk mencapai konvergensi optimal. Proses fine-tuning dilakukan untuk menentukan kombinasi parameter terbaik, seperti yang dijelaskan pada bagian Eksekusi Pelatihan.
- 5) Analisis Hasil Pelatihan:

- a) Program Pertama (Pelatihan Model FER): Mengukur Train Loss dan Test Accuracy per epoch, divisualisasikan melalui kurva loss dan akurasi.
- b) Program Kedua (Fine-Tuning Model Face Detection): Mengukur `box_loss`, `classification loss (cls_loss)`, `distribution focal loss (dfl_loss)`, precision, recall, mean Average Precision (mAP) pada ambang 0.5 (`mAP@50`), dan mAP rata-rata dari ambang 0.5 hingga 0.95 (`mAP@50–95`), dan waktu inferensi (termasuk waktu preprocess dan postprocess per citra) untuk mengevaluasi performa deteksi wajah. Hasil dari fine-tuning disimpan di dalam Google Drive.
- c) Program Ketiga (Fine-Tuning Model FER pada Dataset Classroom): Mengukur Average Train Loss, Validation Loss, Validation mAP, dan Average Precision (AP) untuk masing-masing dari tujuh kelas emosi (Angry, Disgust, Fear, Happy, Neutral, Sad, Surprise).
- d) Program Keempat (Fine-Tuning Model Emotion Detection): Mengukur `box_loss`, `classification loss (cls_loss)`, `distribution focal loss (dfl_loss)`, mean Average Precision (mAP) pada ambang 0.5 (`mAP@50`), dan mAP rata-rata dari ambang 0.5 hingga 0.95 (`mAP@50–95`), baik secara keseluruhan maupun per kategori emosi, dan waktu inferensi (termasuk waktu preprocess dan postprocess per citra) untuk mengevaluasi performa deteksi emosi. Hasil dari fine-tuning disimpan di dalam Google Drive.

4. Evaluasi

Tahap evaluasi bertujuan untuk mengukur performa model pada data uji dan memvalidasi hasil penelitian. Langkah-langkahnya meliputi:

- 1) Analisis Metrik Evaluasi:
 - a) Program Pertama (Testing Model FER): Evaluasi performa klasifikasi pada dataset RAF-DB dilakukan menggunakan confusion matrix, precision, recall, dan F1-score. Selain itu, laporan klasifikasi (classification report) dihasilkan untuk masing-masing kelas emosi guna memberikan gambaran performa model secara per kelas. Proses

inferensi juga dilakukan dengan menguji model pada beberapa gambar wajah dari dataset RAF-DB untuk mengevaluasi akurasi prediksi secara langsung.

- b) Program Kedua (Testing Model Face Detection): Mengevaluasi prediksi bounding box pada gambar validasi, divisualisasikan dengan kotak ground truth (hijau) dan prediksi (biru) untuk memverifikasi akurasi deteksi wajah.
- c) Program Ketiga (Testing Pipeline FER pada Dataset Classroom): Mengukur nilai mean Average Precision (mAP) pada data uji, Average Precision (AP) untuk masing-masing kelas emosi, serta latensi rata-rata per wajah dan per citra (waktu inferensi), disertai dengan penghitungan frame per second (FPS) dari keseluruhan pipeline deteksi wajah dan pengenalan ekspresi wajah. Selain itu, dilakukan visualisasi hasil prediksi untuk dibandingkan dengan data ground truth. Confusion matrix yang telah dinormalisasi digunakan untuk mengevaluasi performa klasifikasi secara lebih rinci pada masing-masing kategori emosi.
- d) Program Keempat (Testing Model Emotion Detection): Mengukur precision, recall, mean Average Precision (mAP) pada ambang IoU 0.5 (mAP@50), serta rata-rata mAP dari ambang IoU 0.5 hingga 0.95 (mAP@50–95), baik secara keseluruhan maupun per kategori emosi. Selain itu, diukur pula waktu inferensi (termasuk proses preprocessing dan postprocessing), latensi total (gabungan dari preprocessing, inference, dan postprocessing), serta nilai FPS per wajah dan per citra untuk menilai efisiensi dan kecepatan sistem deteksi emosi pada dataset uji (Sesi 3). Confusion matrix yang telah dinormalisasi digunakan untuk mengevaluasi performa klasifikasi secara lebih rinci pada masing-masing kategori emosi.
- e) Interpretasi Hasil: Metrik evaluasi dianalisis untuk mengidentifikasi kekuatan dan kelemahan model, serta untuk memastikan bahwa model dapat digeneralisasi pada skenario dunia nyata seperti lingkungan kelas.