PROJECT PROPOSAL

SUPERVISE ACCESS TO TOURIST AREAS AND INDUSTRIAL PARKS; PREDICTING EMOTIONS, GENDER AND AGE.

Group 4: Trần Xuân Thành

Supervisor: Mr. Nguyễn Quốc Trung

Subject: DAT301m

Semester: Spring 2024

Ho Chi Minh City

Table of Contents

I. Introduction

1. Overview

a) *Project Information*

- Project name: *Supervise access to tourist areas and industrial parks; Predicting emotions, gender and age.*
- Group name: *Group 4.*

b) *Project Team*

- Supervisor

Full Name	Email	Mobile	Title
Nguyen Quoc Trung	trungnq46@fe.edu.vn trungnq46@fpt.edu.vn	0979350707	Lecturer

Table 1. *Supervisor contact information*

- Team Members

Full Name	Email	Mobile	Role
Trần Xuân Thành	ThanhTXSE160972@fpt.edu.vn	0389086735	Leader, Member

Table 2. *Team Members contact information*

2. Background and Context

In the current era, Artificial Intelligence is being promoted in many industries and fields in Vietnam, bringing obvious benefits. Serving the implementation of *Official Dispatch No. 4033/TCTTKĐA* dated June 13, 2023 of the Working Group implementing *Project 06 of the Government* on implementing pilot models to promote the implementation of Project 06, *dated November 1 In 2023*, organizations are researching and developing camera models that apply artificial intelligence, including *Model 13: Deploying AI Cameras to control entry and exit to tourist areas* and *Model 14: Deploying AI Cameras to control entry and exit to the*

Industrial Park. With the model of controlling and detecting people loitering in front of industrial parks, there is currently only model of the KBVISION AI camera, but the definition of loitering is too simple. Other models have not achieved high performance.

In this project, we develop a loitering detection model on multiple surveillance cameras with a definition of loitering more suitable to the actual situation in Industrial Parks in Vietnam.

The challenge of this project is the need to monitor on multiple cameras, meaning the subject's ID must be synchronized across the cameras to track and detect the loitering behavior.

In addition, in this project, we further develop a model to count the number of people entering/exiting the tourist area's gates and predicting customers' emotions, gender, and age.

II. Related Works

1. YOLOv8 Detection and Tracking:

YOLOv8 is a state-of-the-art object detection and tracking model that offers significant advantages over previous models in terms of speed, accuracy, and versatility. It is well-suited for real-time applications such as autonomous vehicles and video surveillance.

Compared to other models:

- YOLOv8 is faster and more accurate than YOLOv5, with improved efficiency.
- It is faster, more accurate, and has fewer parameters than SSD.
- YOLOv8 is faster and more efficient than Faster RCNN.

Benefits of using YOLOv8:

- High performance: meets requirements for speed and accuracy.
- Flexibility: applicable to various fields and tasks.
- Easy to deploy: open source, with comprehensive documentation.

Conclusion:

YOLOv8 is an optimal choice for person detection and tracking applications that demand high speed, high accuracy, and flexibility.

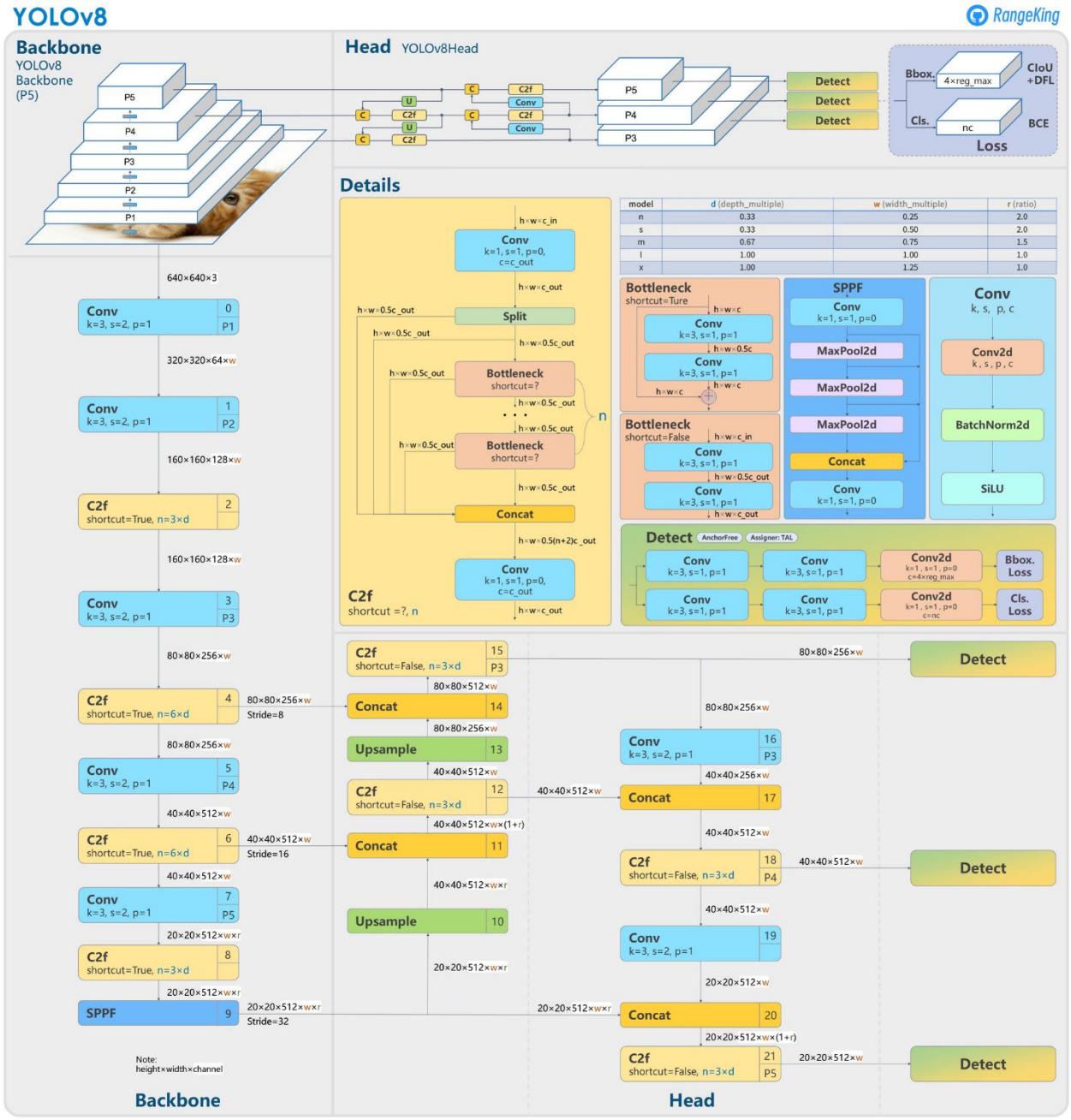

Figure 1. *YOLOv8 Architecture, visualisation made by GitHub user RangeKing*

2. "Age, Gender Prediction and Emotion recognition using Convolutional Neural Network" (by Arjun Singh, Nishant Rai, Prateek Sharma, Preeti Nagrath and Rachna Jain, 2021)

In the research paper *"Age Gender Prediction and Emotion Recognition using Convolutional Neural Network"* by Arjun Singh, Nishant Rai, Prateek Sharma, Preeti Nagrath, and Rachna Jain, the authors present a sophisticated method for predicting age and gender, and recognizing emotions, using Convolutional Neural Networks (CNNs). This study is particularly relevant to the field of visual surveillance and biometric analysis.

Key aspects of the paper include:

The development of two distinct models: one for age and gender prediction utilizing the Wide Residual Network (Wide ResNet) architecture, and another for emotion recognition employing a conventional CNN framework.

The utilization of the IMDb-WIKI dataset for age-gender classification and the Fer2013 dataset for emotion recognition.

A detailed description of both models' architectures, highlighting the Wide ResNet for its "flat" and "wide" variation of the standard ResNet.

Significant results, with the age-gender prediction model achieving an accuracy of 96.26%, and the emotion recognition model reaching 69.2% accuracy.

These findings offer valuable insights for advancements in automated age and gender detection, especially in the presence of varying facial expressions, and contribute to the growing body of knowledge in emotion recognition through neural networks.

III. Project Design

1. Project Solution Overview

- *Aims and Scopes:*
  - ✓ Develop a model to monitor and detect loitering behavior in front of industrial park gates in accordance with the actual situation in Vietnam.
  - ✓ Control the number of people entering and exiting the tourist area gate.
  - ✓ Try to build emotion, gender and age prediction model and improve accuracy of existence one
  - ✓ After that, save the object's information into the database or send a warning to the management department.
- *Workflow:*

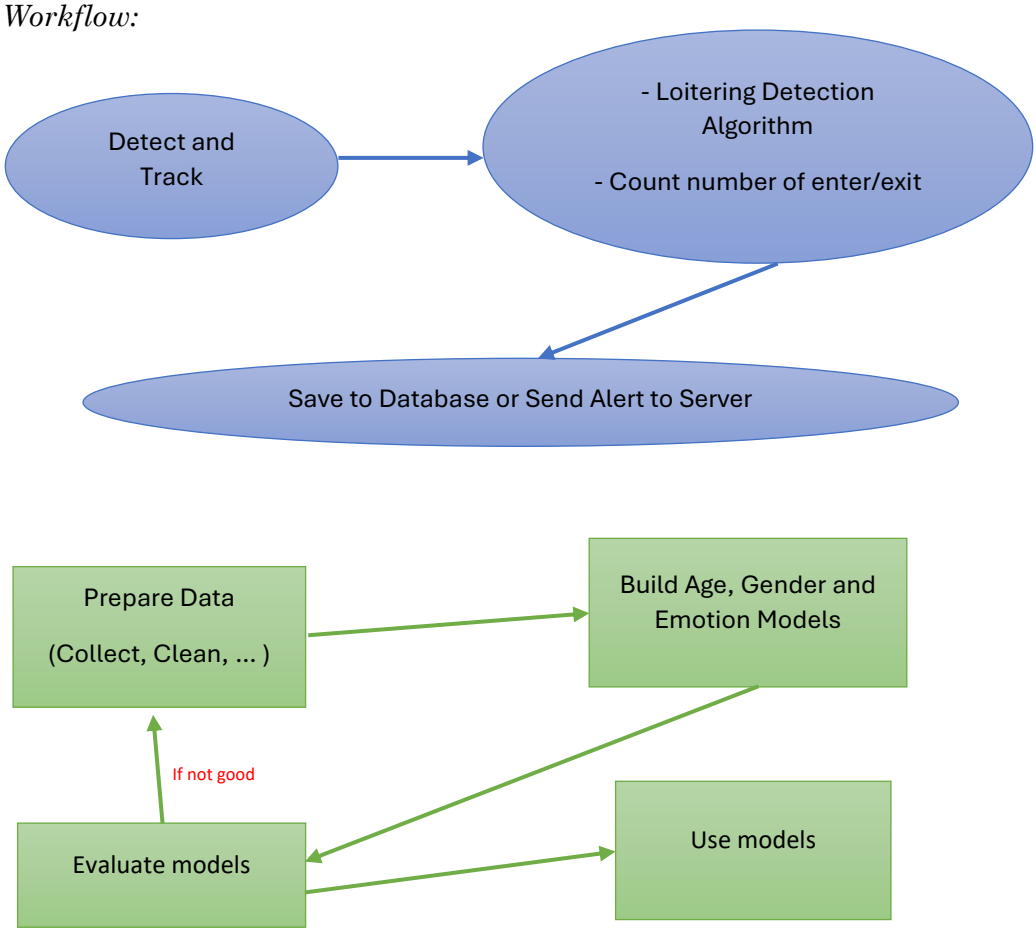

Figure 2. *Project Workflows*

2. Project Materials

a. UTKFace

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. This dataset could be used on a variety of tasks, e.g., face detection, age estimation, age progression/regression, landmark localization, etc. Some sample images are shown as following:

Figure 3. *Some sample images of UTKFace dataset.*

We will use this dataset to train the age and gender model.

b. Emotion Detection (kaggle.com)

The dataset contain 35,685 examples of 48x48 pixel gray scale images of faces divided into train and test dataset. Images are categorized based on the emotion shown in the facial expressions (happiness, neutral, sadness, anger, surprise, disgust, fear).

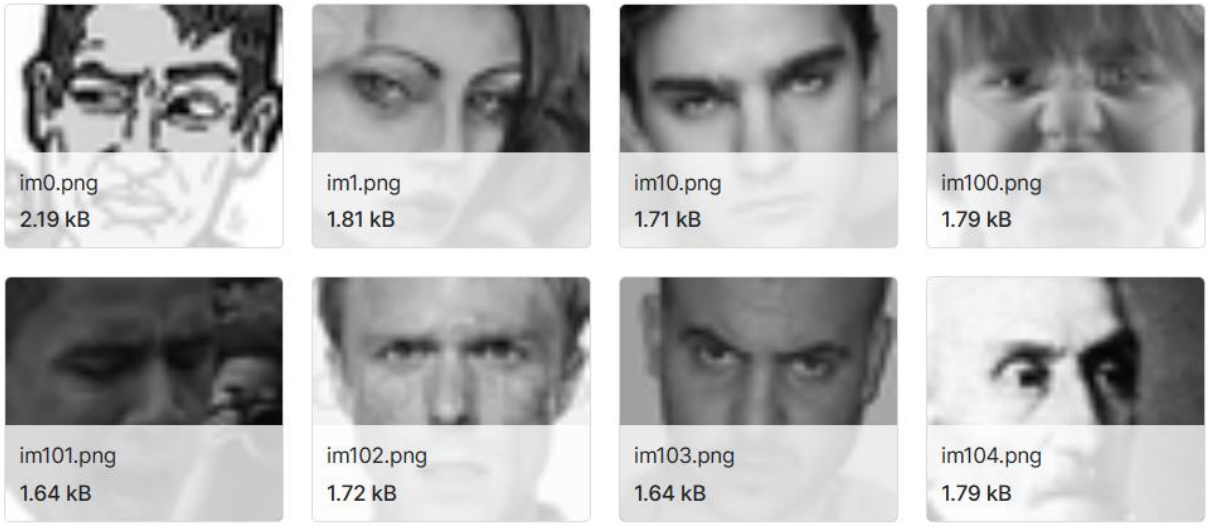

Figure 4. *Some sample images of the dataset*

We will use this dataset to train the Emotion model.

Link to the dataset: *https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer*

3. Project Methods

State	Algorithm	Pretrained ?	Description
People Detection and Tracking	YOLOv8	Yes	Get objects' ID and bounding box.
Loitering Detection	Loitering Detection Algorithm	Our Rules.	- Exceeds the time threshold long enough. - Exceeded the threshold number of changes in direction of movement.
Get faces from images/videos	Haar Cascade	Yes	Get faces' bounding box.
Emotion, Gender and Age Prediction	CNN models	Build from scratch	

Table 3. *Project Methods*

Evaluation Metric

Intersection over Union (IoU): IoU is a measure that quantifies the overlap between a predicted bounding box and a ground truth bounding box. It plays a fundamental role in evaluating the accuracy of object localization.

Average Precision (AP): AP computes the area under the precision-recall curve, providing a single value that encapsulates the model's precision and recall performance.

Mean Average Precision (mAP): mAP extends the concept of AP by calculating the average AP values across multiple object classes. This is useful in multi-class object detection scenarios to provide a comprehensive evaluation of the model's performance.

Precision and Recall: Precision quantifies the proportion of true positives among all positive predictions, assessing the model's capability to avoid false positives. On the other hand, Recall calculates the proportion of true positives among all actual positives, measuring the model's ability to detect all instances of a class.

F1 Score: The F1 Score is the harmonic mean of precision and recall, providing a balanced assessment of a model's performance while considering both false positives and false negatives.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

TP = True positive

TN = True negative

FP = False positive

FN = False negative

Figure 5. Precision and Recall formulas

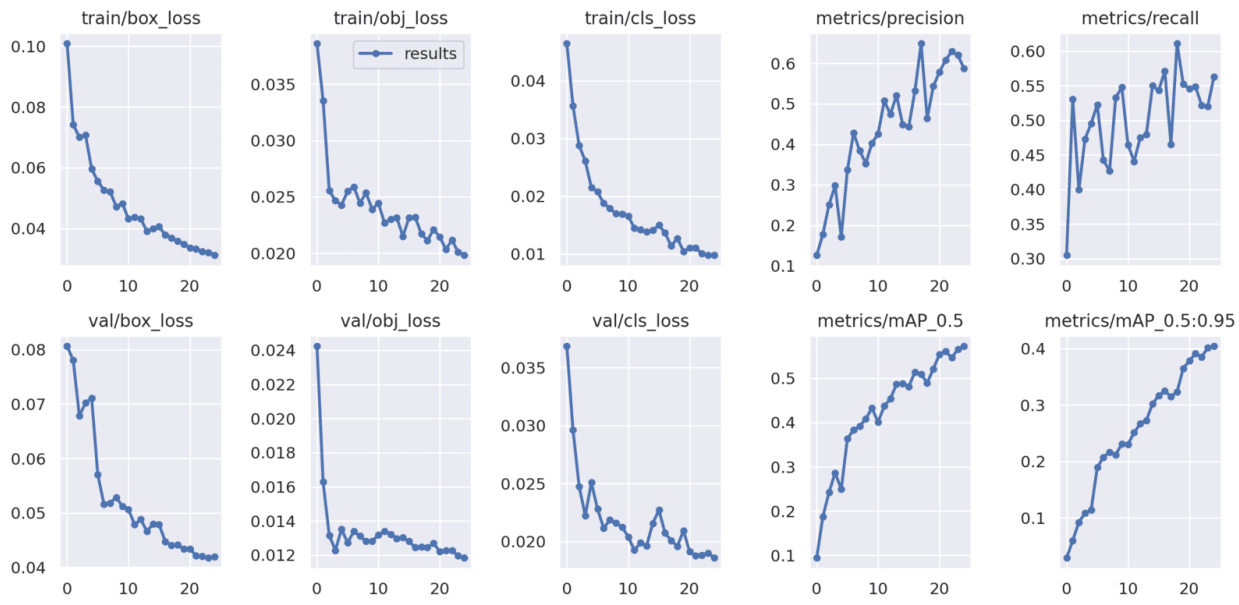

Figure 6. *Evaluation Metrics of YOLOv8 models*

IV. Project Management

1. Project Schedule

Project Works	Timeframe
Planning	Deadline: 15/1
Learn about the YOLOv8 model	15/1 – 22/1
Implement loitering detection and people counting algorithms.	22/1 – 29/1
Prepare data to build emotion, gender and age models	15/2 – 22/2
Build and evaluate models	22/2 – 29/2
Execute and complete the project	1/3 – 7/3

Table 4. *Project schedules*

2. Project Resource Management

Platform	Description	Link
Spring Success blog – My Notes *Category:* DAT Project	- Task management - Resource management	*Spring Success Notes*
Google Drive	- Dataset - Backup resource	Link Drive
GitHub	- Code management - Final publish project	thanhtranfpt (Trần Thành) (github.com)

Table 5. *Project management platforms*

REFERENCES

[1] *https://paperswithcode.com/task/age-estimation*

[2] *https://blog.roboflow.com/whats-new-in-yolov8/*

[3] *https://youtu.be/yWZLu7E50x0*

[4] *https://www.kaggle.com/code/springsuccess/age-prediction-from-images-cnn-regression*

[5] *https://docs.ultralytics.com/guides/yolo-performance-metrics/#introduction*

[6] *Arjun Singh, Nishant Rai, Prateek Sharma, Preeti Nagrath and Rachna Jain. "Age, Gender Prediction and Emotion recognition using Convolutional Neural Network", 2021.*