Interim Evaluation Report for I&T Trial Project

Implementation Service for Artificial Intelligence Video Analytic System for Escalator Monitoring at Market

I&T Project No. : P-0066
I&T Wish No. : W-0154
I&T Solution No. : S-0025

Copyright Notice And Disclaimer

Disclaimer

The information contained in this report is compiled by the Electrical and Mechanical Services Department of the Government of the Hong Kong Special Administrative Region ("the Government") for general information only. Whilst the Government endeavours to ensure the accuracy of this general information, no statement, representation, warranty or guarantee, express or implied, is given as to its accuracy or appropriateness for use in any particular circumstances.

This report can also contain information contributed by others over whom, and in respect of which, the Government may have no influence.

The Government is not responsible for any loss or damage whatsoever arising out of or in connection with any information in this report. The Government reserves the right to omit, suspend or edit all information compiled by the Government in this report at any time in its absolute discretion without giving any reason or prior notice. Users are responsible for making their own assessment of all information contained in this report and are advised to verify such information by making reference, for example, to original publications and obtaining independent advice before acting upon it.

Table of Contents

Purpose of the Project and Target Deliverables 4

Project Description 4

Trial Site 5

Trial Timeframe 6

Name and Background of I&T Solution Provider 6

Details of Implemented Trials 7

I. Terminology 7

II. Methodology and Applicable Standards 9

III. User Interface 9

Interim Review on Performance and Outcomes 11

Reference and Appendices 12

Purpose of the Project and Target Deliverables

The project is to design, develop, delivery and install Video Analytics Applications and support and maintenance services for escalator monitoring in wet market. This video analytic system is developed and deployed at market to monitor the operation and safety of escalators. These include misuse of passengers, e.g. carrying of bulky objects to use the escalator, accident detection, e.g. people tripping in the escalator, misconduct of maintenance, e.g. floor cleaning to cause over-wetting the escalator plates etc.

Project Description

Total six cameras are installed at the escalators between the ground floor and the first floor of the Market. These installed cameras together with artificial intelligence (AI) video analytic computing system will be used to provide the following functions:

(a) Face redaction function which is used to mask out all faces of people in the recorded video to prevent individual person from being identified.

(b) AI video analytic services to detect the following events

a. Bulky object detection

b. Accident detection

c. Escalator On/Off and speed status

d. Over-wetting of escalator plates

If the above events are detected, email alerts will be sent out and the videos of the event will be uploaded to EMSD Cloud. Also, a monthly report will be generated to summarize the occurrences of these events in each month.

Trial Site

All six cameras are installed to monitor the escalators. The AI video analytic system (AI supercomputer, Email bridge computer, NAS, UPS etc.) is installed at the management office of the Market.

The system architecture of the AI video analytic system is shown as follows.

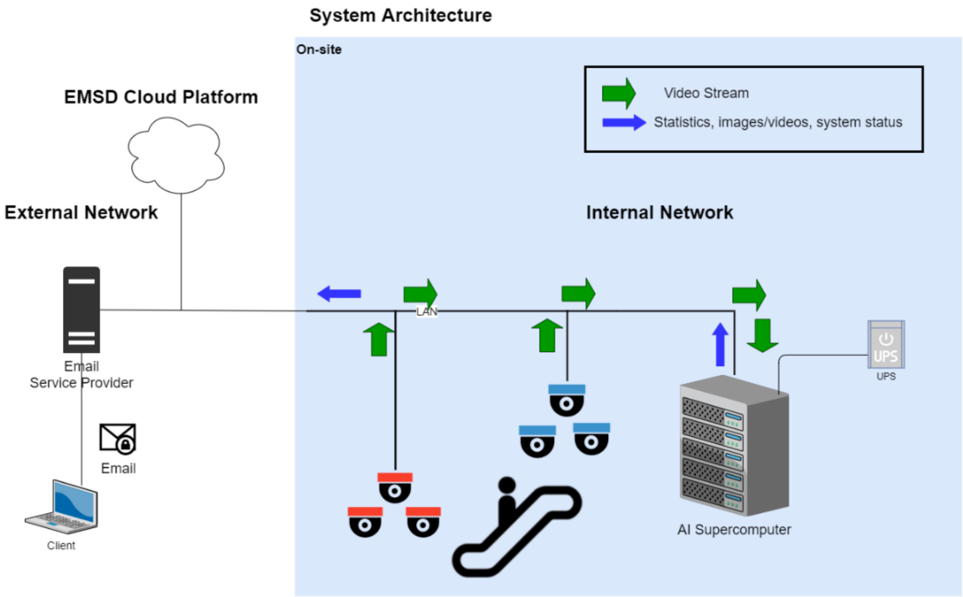


Figure 1. System Architecture of the AI video analytic system

The major equipment of this system are six network cameras, AI supercomputer, email bridge, network storage and equipment, plus un-interruptible power supply.

Trial Timeframe

Time	Details
End / Sept 2019	Project kick off
Oct ~ Dec 2019	Site data collection or simulation
Dec 2019	On-site installation of equipment and cable wiring
30 Dec 2019	Preliminary test on site in non-office hours
Jan ~ March 2020	Validation with on-site acting and fine tune

Name and Background of I&T Solution Provider

This AI video analytic system is wholly developed in-house by Sebit Company Limited.

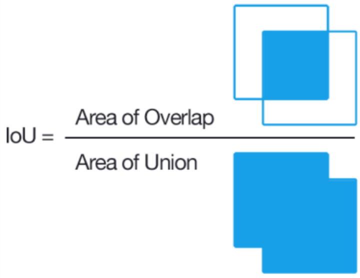
Sebit is a Hong Kong technology company focusing on Artificial Intelligence, Computer Vision and 3D visualization.

Details of Implemented Trials

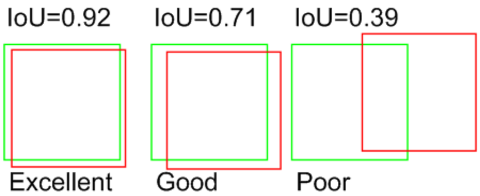
I. Terminology

a. Intersection over Union (IoU)

IoU (Intersection over union) measures the overlap between 2 boundaries. We use IoU to measure how much our predicted boundary overlaps with the ground truth (the real object boundary).



The IoU value is between 0 to 1, the higher the better.



When we evaluate the accuracy of a model, we evaluate it against a predefine IoU threshold (typically 0.5). A detection is considered as "correct" if the output bounding box overlap with the group truth bounding box, with an IoU >= 0.5.

b. Average Precision (AP)

Mathematically, average precision is the area under the plot of precision-recall curve:

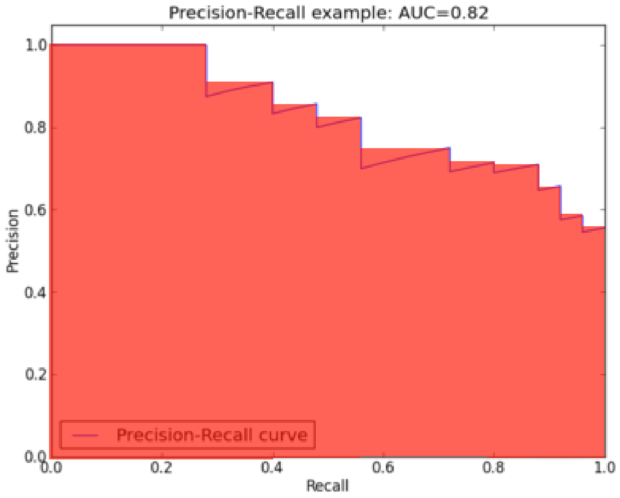
$$AP = \int_0^1 p(r)dr$$

where

$$Precision(p) = \frac{tp}{tp + fp}$$

$$Recall(r) = \frac{tp}{tp + fn}$$

True Positive (tp): number of correct predictions on positive result
True Negative (tn): number of correct predictions on negative result
False Positive (fp): number of incorrect predictions on positive result
False Negative (fn): number of incorrect predictions on negative result



The value of average precision is between 0 and 1, the higher the better. Sometimes, people express average precision in %, from 0% to 100%.

c. Mean Average Precision (mAP)

When it comes to multiclass object detection, each class has its own AP. Hence we take the mean of all the AP to give rise to mean Average Precision (mAP) which is a single numeric value that represent the accuracy of the model.

For examples, if given a model that detects baby strollers with AP = 0.5, wheelchairs with AP = 0.2 and delivery trolleys AP = 0.8. Then the mAP of this model is (0.5 + 0.2 + 0.8) / 3 = 0.5 , or 50%

Because of its simplicity, mAP is a popular metric use to measure the accuracy of object detectors such as RCNN, SSD, YOLO as shown below:

Model	Input size	Train set	Test set	mAP	FPS
YOLOv1	448x448	VOC 2007+2012	VOC 2007	63.4%	45
Fast YOLOv1	448x448	VOC 2007+2012	VOC 2007	52.7%	155
YOLOv2	416x416	VOC 2007+2012	VOC 2007	76.8%	67
tiny-YOLOv2	416x416	VOC 2007+2012	VOC 2007	57.1%	**207**
YOLOv2	608x608	COCO	COCO	48.1%	40
YOLOv3	608x608	COCO	COCO	**57.9%**	20

In this project, the metric we use to evaluate and validate our YoloV3 model is mAP.

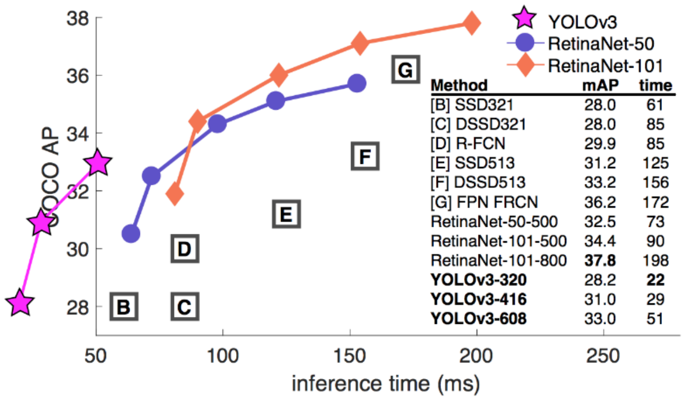
d. Mean Average Precision on IoU (mAP-50, mAP-75, …)

Applying the concept of IoU threshold on mAP, sometimes researchers compares the model performance across different IoU thresholds, that give rise to mAP-50, mAP-70, etc.

II. Methodology and Applicable Standards

For face redaction, YOLOv3, state-of-the-art AI object detection model, is trained to find all human faces in a video, and then mask the faces before the video is saved.

The reason why YOLOv3 is chosen is because of its high speed and accuracy. Below is a graph comparing the performance of different object detection models. In the graph, Y-axis is the accuracy (in mAP), X-axis is the time taken to compute each detection (in ms). It shows that YOLOv3 is much faster than other detection models, but at the same time produce a reasonably accuracy detection mAP.



III. User Interface

Currently the detection events & videos will be uploaded to EMSD cloud and the results are summarized in 3 pages: Events, ON/OFF and Alerts. By clicking into the dots on the timeline, the corresponding video can be viewed.

Figure 10. ON/OFF status

Alert

Logout

Time	Category	Video Capture
20200115-165800	bulky object cam0	Click Here
20200115-165800	bulky object cam2	Click Here
20200115-165800	bulky object cam5	Click Here
20200115-170400	bulky object cam0	Click Here
20200115-170600	bulky object cam0	Click Here
20200115-170600	bulky object cam2	Click Here
20200115-171200	bulky object cam0	Click Here
20200115-171200	bulky object cam2	Click Here
20200115-172000	bulky object cam2	Click Here

Figure 11. Alerts

Interim Review on Performance and Outcomes

After successful installation of AI video analytic system at the market on 16 December 2019, video streams from all six cameras can be face-redacted and recorded at NAS accordingly.

Also, the ON/OFF status and the occurrence of bulky objects are reported on the GUI and corresponding videos were uploaded to EMSD Cloud.

So far in the current model training, we did NOT include any synthetic data, because for the target objects that we need to detect, we manage to collect sufficient & good quality real data from on-site. We anticipate synthetic data may come into place in next stage when we develop the accident detection, because accident video samples are hard to collect.

To further refine the model, the key is to collect more data so that the training samples truly cover and represent the real data distribution. Hence, for model refinement we will perform the following repeatedly: (1) collecting more variety of real data from production (2) review the false positive detections and add those false positive data into the training set (3) deploy the newly trained model

We anticipate the accuracy and functionality of the system will be further improved going forward with more data being collected and fine-tuning.

- END OF REPORT -

Digitalisation and Technology Division

Electrical and Mechanical Services Department

12 March 2020

Reference and Appendices

[1] Redmon, J., Farhadi, A. (2018) 'YOLOv3: An Incremental Improvement'. Available at https://arxiv.org/abs/1804.02767

[2] YOLO website: https://pjreddie.com/darknet/yolo/