Avoiding Stepping Accidents by Computer-vision-based Pedestrian Detection and Prediction

Ai Ran, ChangYan Shih, Jinghao Liang

SIGS, Tsinghua University, China

ABSTRACT

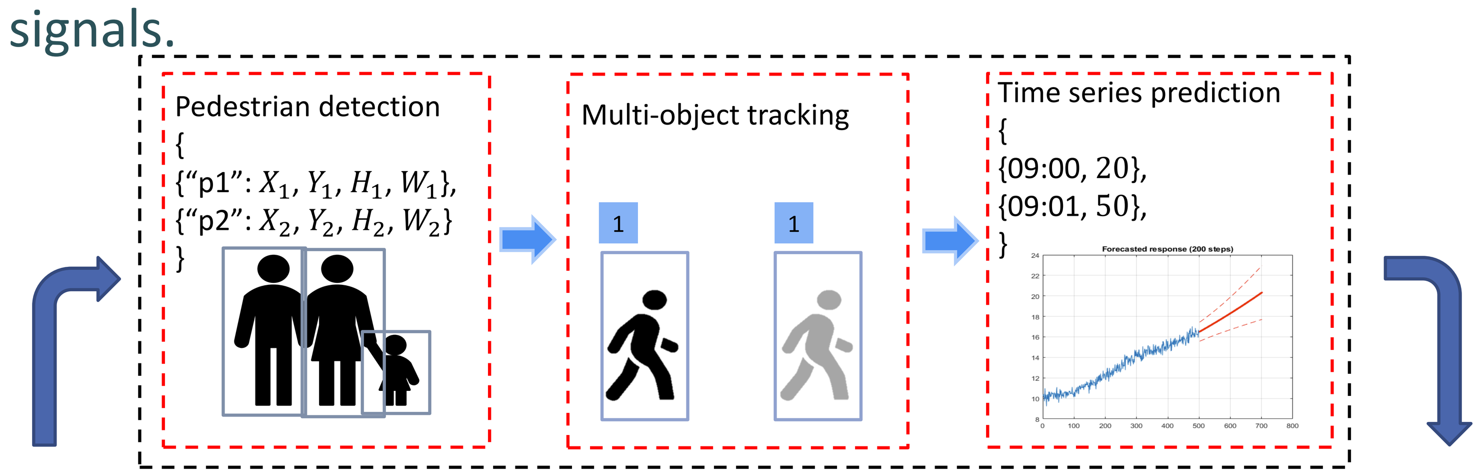
Stepping accidents often happen when a flock of people with large density passes a small and relatively close area. Existing overcrowded hazard detection methods mainly use overhead head detection at the entrances to estimate the population within the area. However, it is unable to detect population flows and difficult to give immediate warnings when accidents are about to occur. We propose a concrete framework to predict stepping accidents via computer-vision (CV) and time series prediction approaches. Empirical experiment results show that our approach can reach a high prediction accuracy.

INTRODUCTION

When accidents come up at the front of a dense unorganized population and stop, the back keeps moving forward and crashes with the front. Some individuals may stumble, and bring the others down, like dominos. When there is not enough space for the crowds to disperse, a stepping outbreaks.

Stepping accidents avoiding system :

- Perform a CV algorithm to identify and count the pedestrians in the whole area.
- Predict the population in future minutes by LSTM.
- When the expected quantity in future periods exceeds a threshold, a warning is sent to all pedestrians by audio broadcasting or visible signals.



REFERENCES

[1] "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,", 2022
[2] "SSD: Single Shot MultiBox Detector,", 2015
[3] "Simple Online and Realtime Tracking with a Deep Association Metric.", 2017
[4] "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network", 2020
[5] "Understanding LSTM Networks", http://colah.github.io/posts/2015-08-Understanding-LSTMs/
[6] "xgboost: eXtreme Gradient Boosting", 2017.

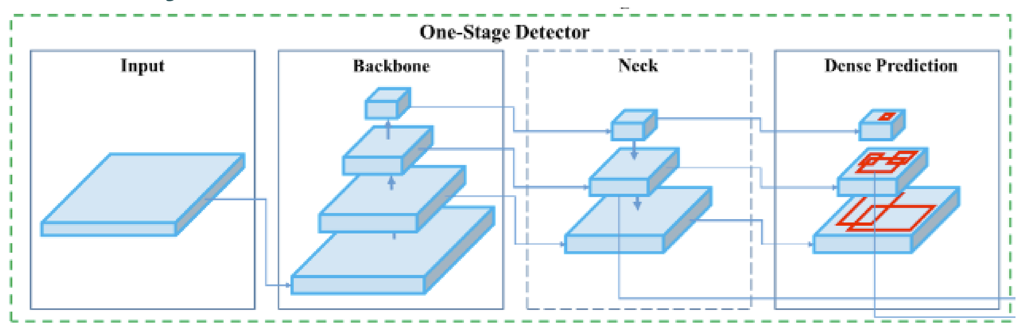
METHOD

Dataset: Wildtrack

- 7 series of 400 annotated frames in a square, 2 frames per second.

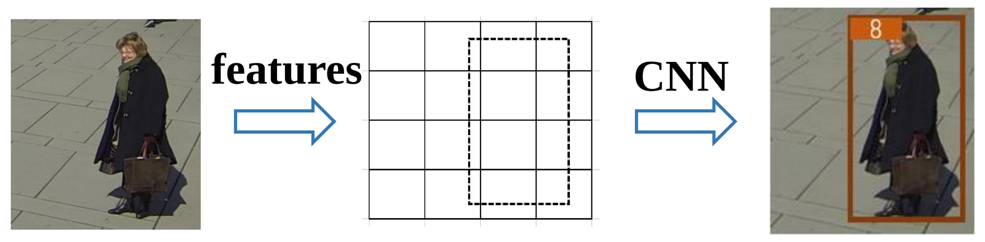
You Only Look Once (YOLOv7)[1]

1. Extract features through a backbone.
2. Aggregate features in the neck.
3. Passed along to the head of the network to predicts the locations and classes of objects.



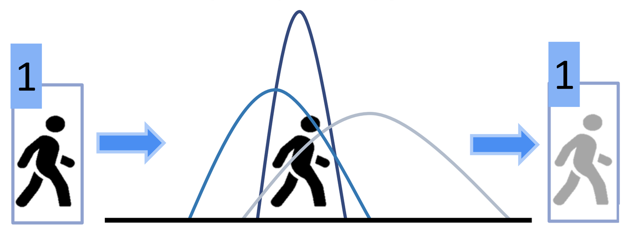
Single Shot MultiBox Detector (SSD)[2]

1. Find bounding box using Mobile-Net V3 + SSD.
2. Extract features(128 dim) detected in step 1 using a CNN.

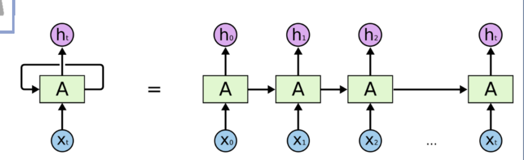


Deep Simple Online Realtime Tracking (DeepSORT)[3]

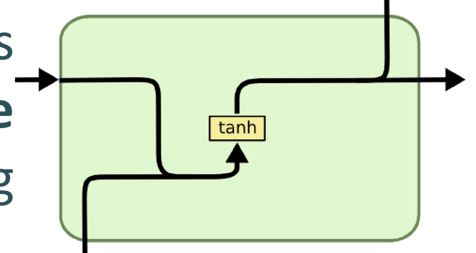
1. Extract features from the detected objects and create a feature embedding for each object.
2. Use the feature embeddings to assign unique IDs to each object.
3. Use the Kalman filter to predict the location of each object in the next frame.



Long Short Term Memory (LSTM)[5]



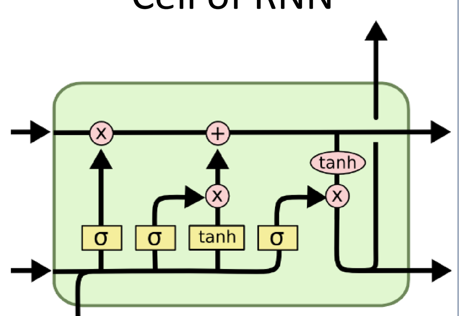
Recurrent Neural Networks (RNN)[4] performs well in handling data series. Weights of the previous cell are passed to the next cell. Instead of only **one** previous cell, LSTM advances RNN in having selective memories from **multiple** previous cells.



Cell of RNN

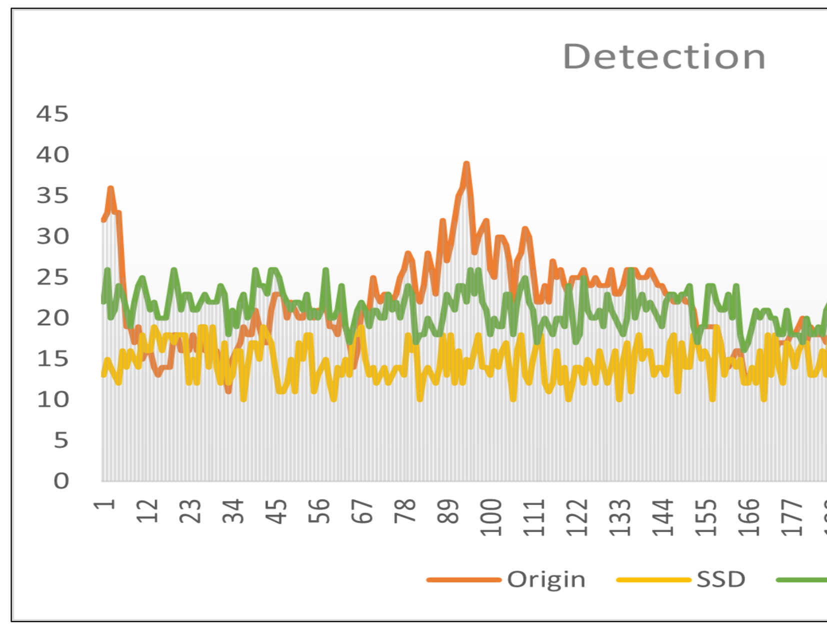
eXtreme Gradient Boosting (XGB)[6]

XGB extracts features in data series by decision trees. In each iteration, the decision tree is updated by the residuals of the previous trees. Features extracted from the series are then applied for regressions to predict future entities.



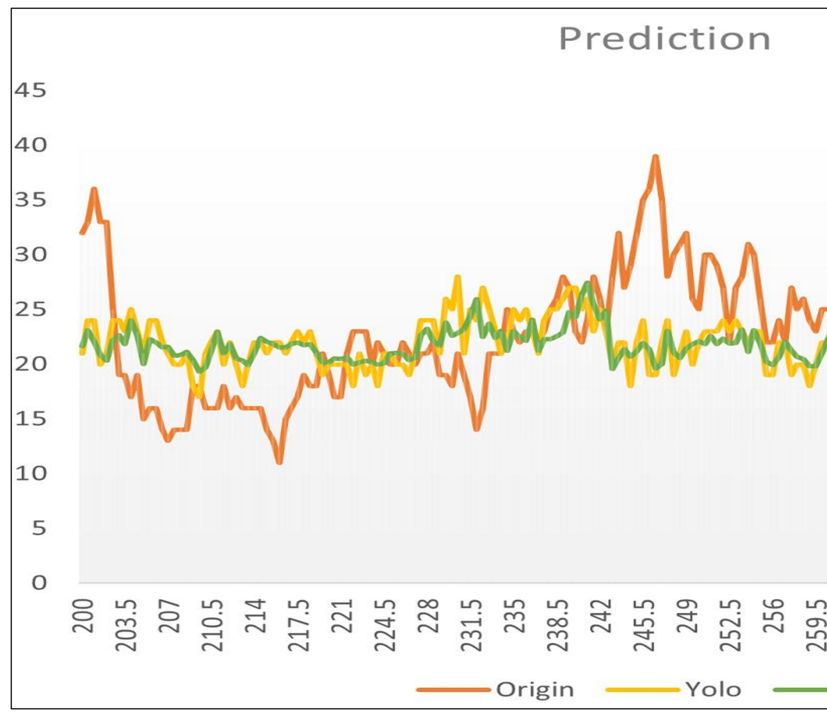
Cell of LSTM

RESULTS



Detection	RMSE	MAPE
YOLO	5.523	21.541%
SSD	9.275	31.427%

The number of people detected is less than the original data. Both detection algorithms did well at detecting clear object at front, while have trouble identify those at the back.



Prediction	RMSE	MAPE
LSTM	1.484	5.250%
XGB	2.254	8.620%

We perform LSTM and XGB on the results from YOLO. Due to the limitation of sample sizes, LSTM outperforms XGB. LSTM predicts YOLO sequence well.

CONCLUSIONS

- New image label strategy and data augmentation are required to get better accuracy.
- Adjust the monitor angle to show more significant feature from the pedestrians to achieve the original goal with same model.
- With the above methods to improve the accuracy of the detection model, the time series model can be used to predict the number of people in the future, so as to effectively prevent stampede incidents.