Multiple Object Forecasting: Predicting Future Object Locations in Diverse Environments

Olly Styles, Tanaya Guha, and Victor Sanchez

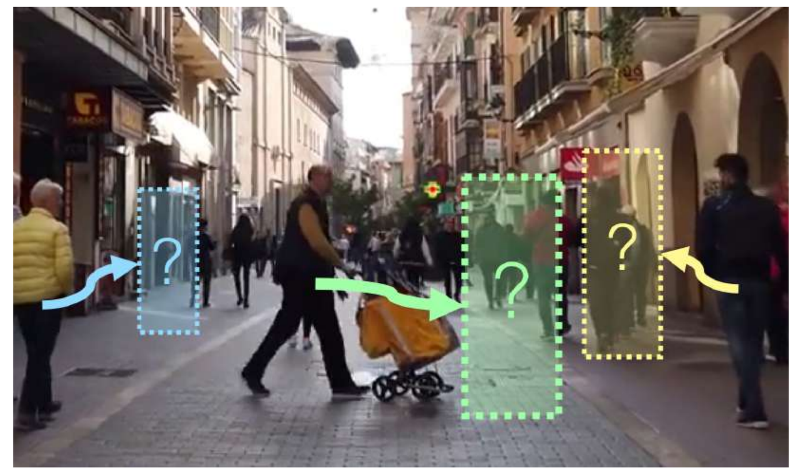
Department of Computer Science, University of Warwick, UK

WACV 2020

Overview

Abstract

- We extend multiple object *tracking* to multiple object *forecasting*
- Given the past 1 second of person bounding boxes, we predict the future 2 seconds of bounding boxes.



Our contributions

1. **New problem**: We introduce Multiple Object Forecasting (MOF), a new formulation of the trajectory forecasting problem.
2. **New dataset**: We introduce Citywalks, a challenging dataset for MOF with considerably more variety than existing datasets.
3. **New model**: We propose STED, a Spatio-Temporal Encoder-Decoder model for MOF which combines visual and temporal features.

References

[1] Redmon, J. and Farhadi, A. Yolov3: An incremental improvement. ArXiv, 2018

[2] He, K., Gkioxari, G., Dollár, P. and Girshick, R. Mask R-CNN. ICCV, 2017

[3] Yagi, T., Mangalam, K., Yonetani, R. and Sato, Y. Future person localization in first-person videos. CVPR, 2018

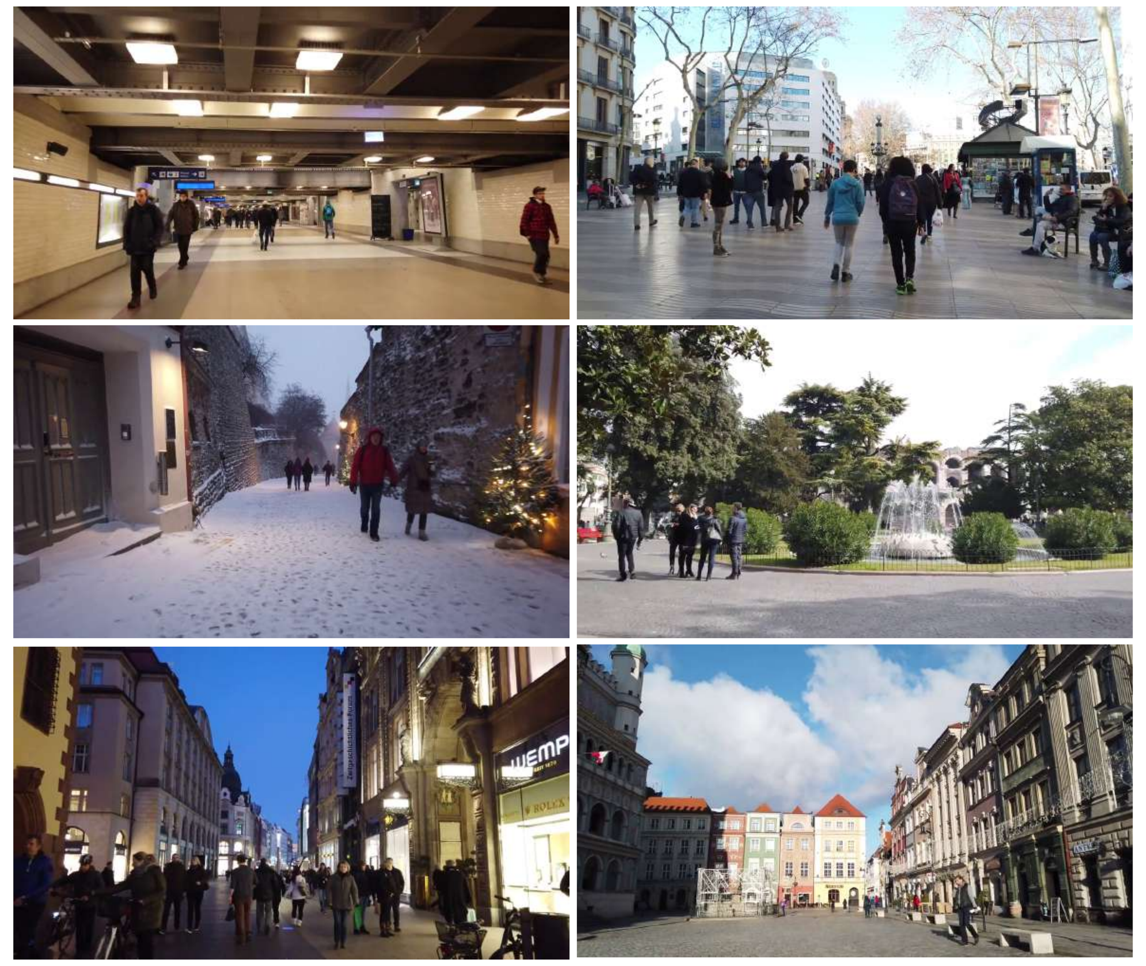
[4] Styles, O., Ross, A. and Sanchez, V. Forecasting Pedestrian Trajectory with Machine-Annotated Training Data. Intelligent Vehicles Symposium, 2019

GitHub *github.com/olly-styles/Multiple-Object-Forecasting*

YouTube *youtu.be/GPdNKE6fq6U* *o.c.styles@warwick.ac.uk*

Details

Our dataset: Citywalks



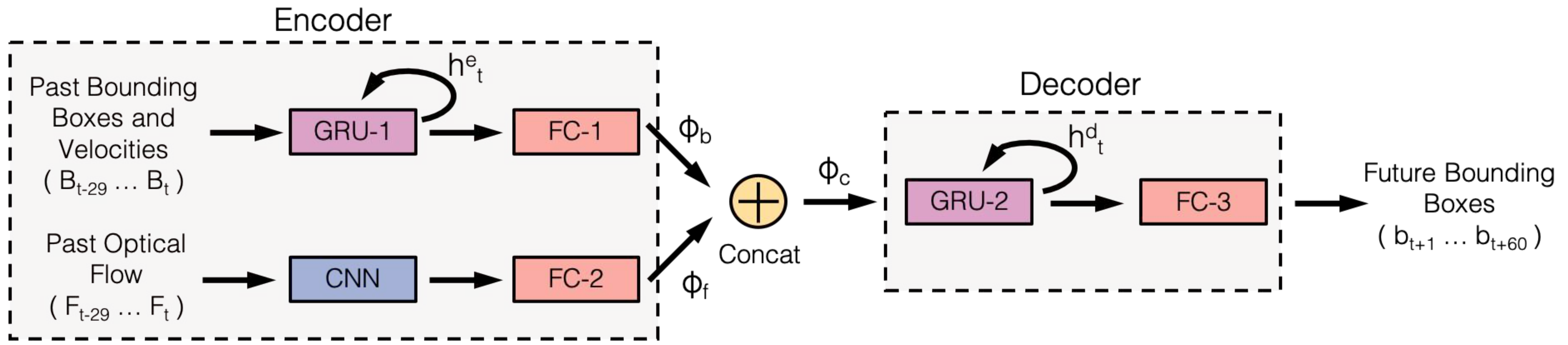
- The Citywalks dataset for Multiple Object Forecasting consists of footage from a handheld camera.
- We use the result of object detection and tracking algorithms as ground truth.

Dataset statistics

Video clips	358
Resolution	1280 × 720
Frame rate	30hz
Clip length	20 seconds
Unique pedestrian tracks	3623
Unique cities	21
Object Detectors	YOLO [1] & Mask-RCNN [2]

Our model: STED

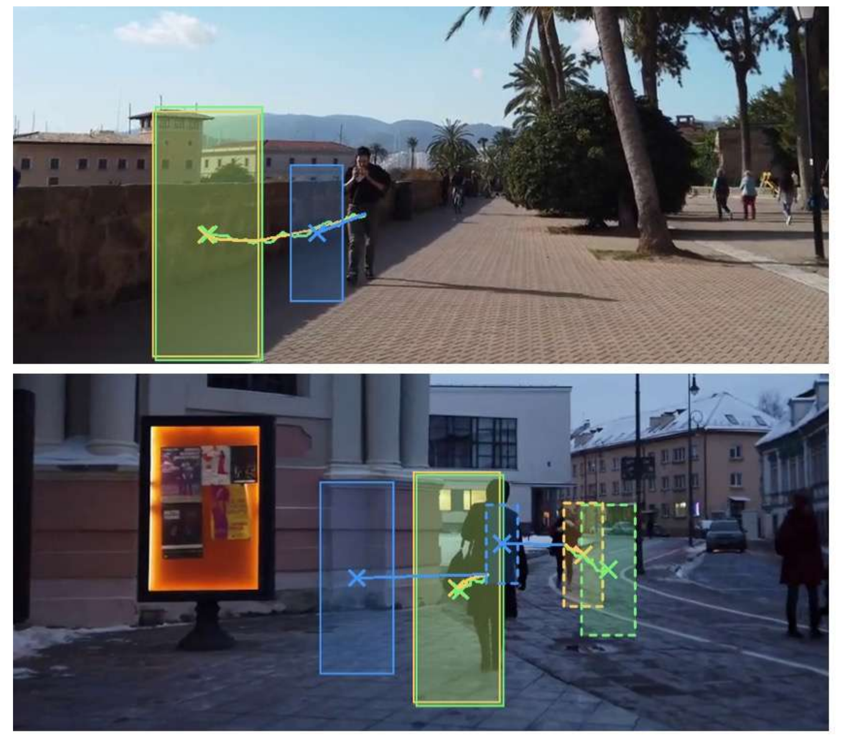
- Our Spatio-Temporal Encoder-Decoder (STED) model uses a GRU and CNN to extract features from past bounding boxes and optical flow, and another GRU for prediction.
- Our experiments show the extracted features are complementary, and STED outperforms prior state-of-the-art methods adapted for Multiple Object Forecasting.



Results

Qualitative results

- Visualization of ground truth, constant velocity/scale, and STED bounding box prediction.
- STED anticipates non-linear changes in velocity and scale.



Quantitative results

Model	YOLOv3		Mask-RCNN	
	ADE/FDE ↓	AIOU/FIOU ↑	ADE/FDE ↓	AIOU/FIOU ↑
Constant velocity	32.9/60.5	51.4/26.7	31.6/57.6	46.0/21.3
Linear Kalman filter	34.3/62.1	49.1/25.5	32.9/59.0	43.9/20.1
DTP [12]	28.7/52.4	-/-	26.7/48.5	-/-
FPL [4]	30.2/53.4	-/-	28.6/49.8	-/-
DTP-MOF [12]	29.0/52.2	54.6/30.8	27.3/49.2	49.6/25.1
FPL-MOF [4]	31.6/55.7	53.0/30.9	29.3/51.0	44.9/22.6
STED	**27.4/49.8**	**56.8/32.9**	**26.0/46.9**	**51.8/27.5**

- STED outperforms existing methods in terms of Average Displacement Error (ADE), Final Displacement Error (FDE), Average Intersection Over Union (AIOU), and Final Intersection Over Union (FIOU)

Acknowledgements

This work is funded by the UK EPSRC (grant no. EP/L016400/1) and the EU Horizon 2020 project IDENTITY (Project No. 690907). Our thanks to NVIDIA for supporting this research with their generous hardware donation.