Forecasting Pedestrian Trajectory with Machine-Annotated Training Data

Olly Styles[1], Arun Ross[2] and Victor Sanchez[1]

[1] Department of Computer Science, University of Warwick, UK
[2] Department of Computer Science and Engineering, Michigan State University, USA

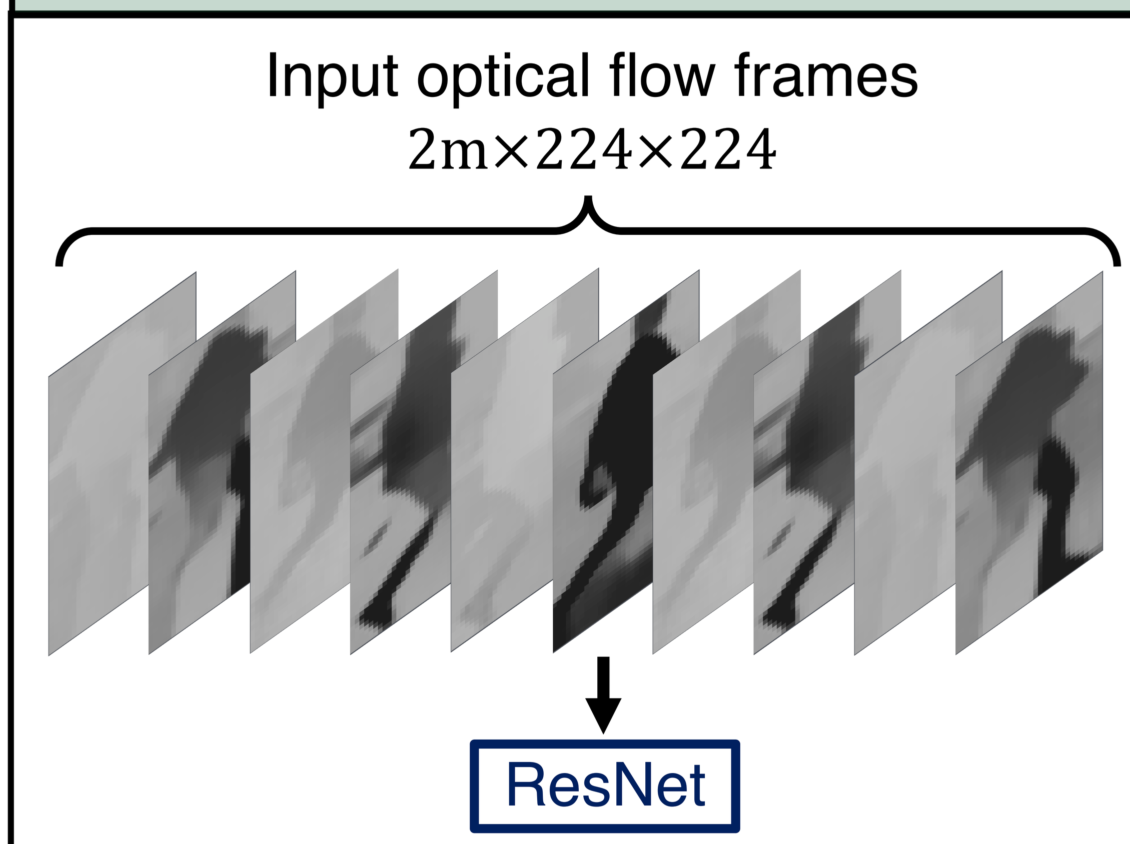
Objectives

1) Predict the future location of pedestrians from onboard a moving vehicle.
2) Learn from unlabeled data.

Contributions

1) Dynamic Trajectory Predictor (DTP), a pedestrian trajectory forecasting deep learning model based on motion features from optical flow.
2) A machine annotation scheme for training trajectory forecasting models in the absence of labeled data.

Our model (DTP)



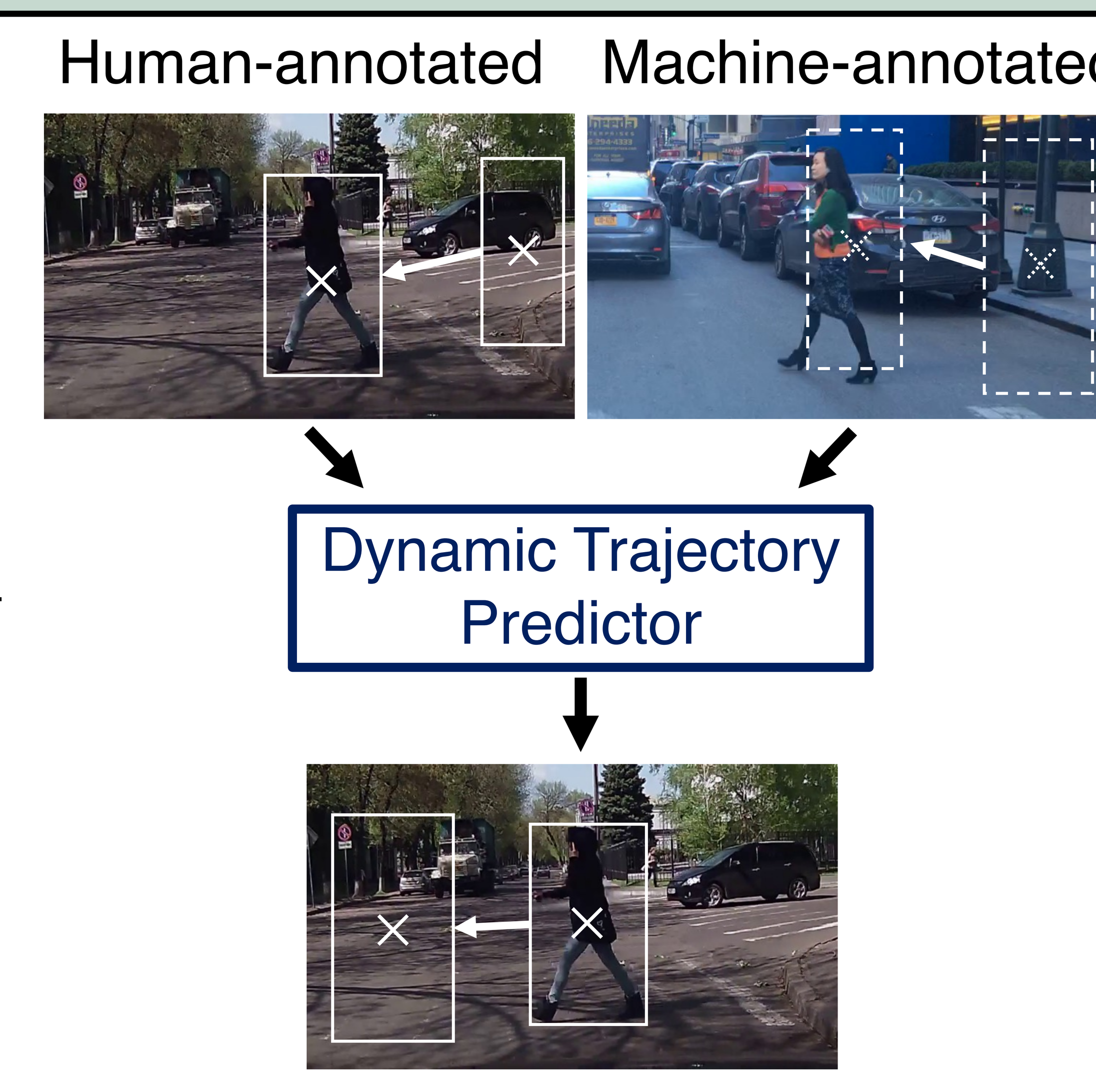
$$L_t + v_t \cdot n = \tilde{L}_{t+n} \qquad \hat{C}_{t+n}$$

$$\tilde{L}_{t+n} + \hat{C}_{t+n} = \hat{L}_{t+n}$$

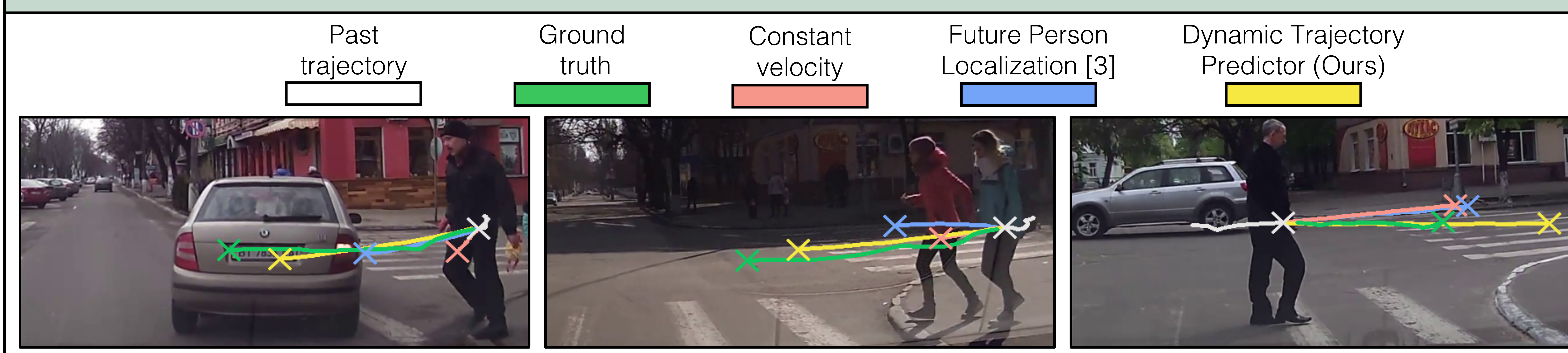
- Optical flow frames extracted from pedestrian bounding boxes.
- CNN (ResNet18) extracts features.
- Predicts a compensation term $\hat{C}_{t+n}$ to correct for errors in the constant velocity equation $L_t + v_t \cdot n = \tilde{L}_{t+n}$

Our annotation scheme

- Data is annotated using a pedestrian detection [1] and tracking [2] algorithm.
- Model is trained using both machine-annotated and human-annotated data.
- Model is evaluated using human-annotated data.



Results



Model	Mean squared error	Final displacement error
Constant acceleration	1426	52.8
Constant velocity	1148	47.5
FPL [3]	881 ± 44	41.3 ± 1.2
DTP (Ours)	**610 ± 21**	**34.6 ± 0.5**
FPL w/ pretrain [3]	805 ± 46	40.1 ± 1.2
DTP w/ pretrain (Ours)	**539 ± 13**	**32.7 ± 0.4**

- DTP performs particularly well in situations where a pedestrian first begins walking (left and centre figures).
- DTP performs less well under conditions of significant background motion (right figure).
- Machine-annotated pretraining improves performance of both DTP and FPL [3].

Conclusion

- DTP forecasts pedestrian trajectory up to 1 second into the future.
- Anticipates non-linear trajectories using optical flow information.
- Using both human-annotated and machine-annotated data improves performance.

References

[1] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. arXiv:1804.02767, 2018.
[2] N. Wojke, A. Bewley, and D. Paulus. Simple online and realtime tracking with a deep association metric. In ICIP, 2017.
[3] T. Yagi, K. Mangalam, R. Yonetani, and Y. Sato. Future person localization in first-person videos. In CVPR, 2018.

Acknowledgement

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