

Ground-truthing Large Human Behavior Monitoring Datasets

Tehreem Qasim, Robert B. Fisher, Naeem Bhatti

Introduction

- Ground-truthing video data is time-consuming
- We exploit: modern detectors and behavior classifiers, plus specialized consistency checks
- Applied to 1 FPS video of small numbers of people
- Reduced labeler clicks by 99+%
- Applicable to behavior monitoring



Office dataset: Day 2, Frames 598-602

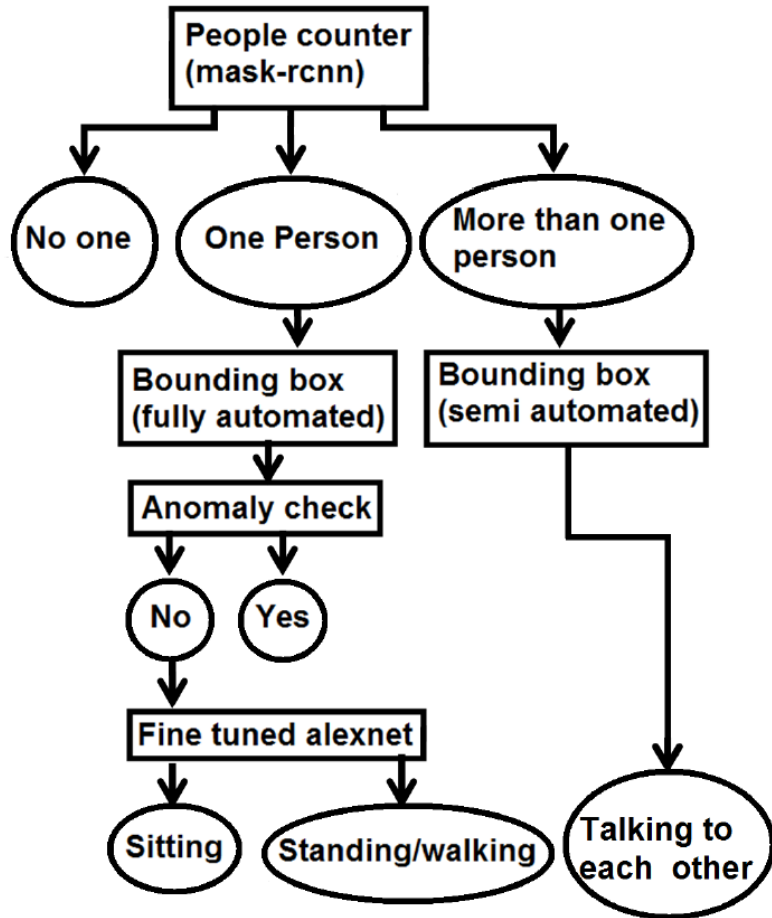
Introduction

Overview of method

Experimental results

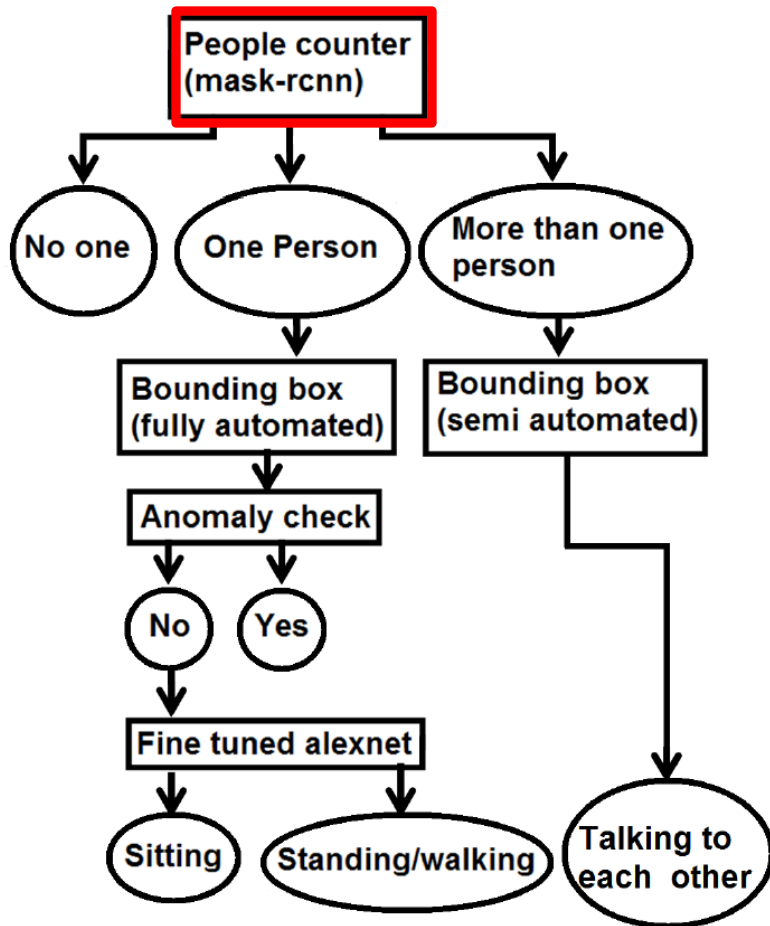
Conclusions

Overview of the method



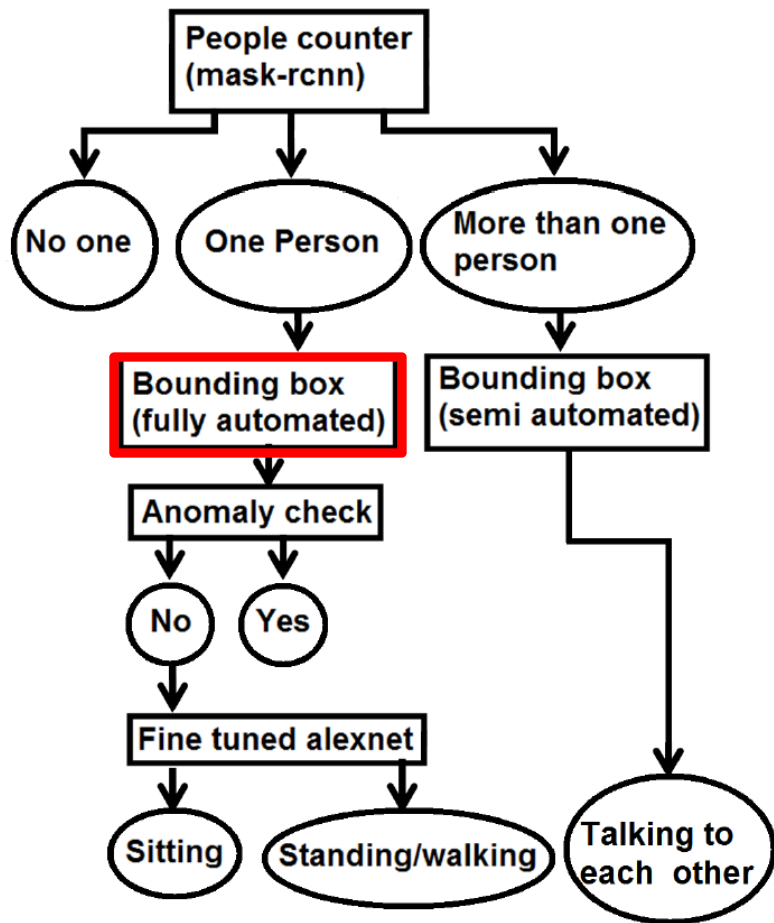
- Mask-rcnn is used as a people counter
- Motion based checks to correct mask-rcnn errors
- Bounding box extraction based on count of people in a frame
- Anomaly detection based on inactivity analysis
- Behavior labeling using a fine-tuned alexnet

People counter



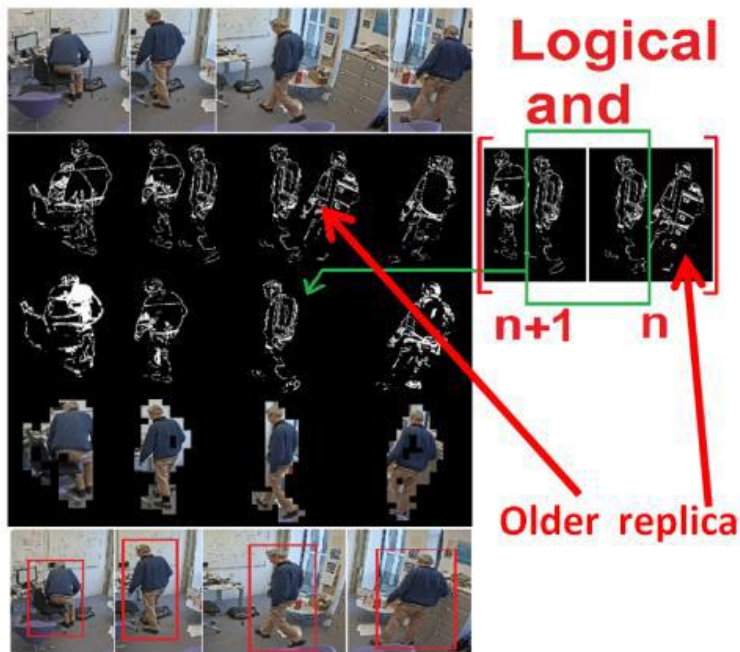
- Initial count of people in video frames obtained from mask-rcnn
- Motion based semi-automated checks to detect and correct errors
- Special tests to correct errors or flag for manual intervention

Single person bounding box method



- For single person frames motion (difference of frame gradient) based method to draw bounding boxes

Single person bounding box method



Older replica problem

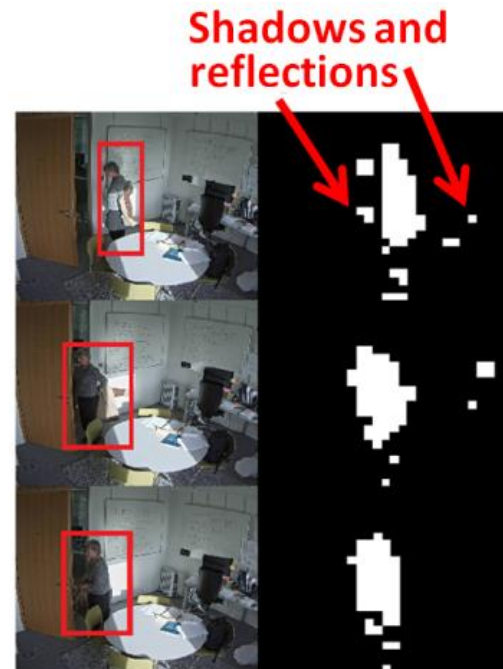
Top row: color frames (day 4, frames 1603-1606)

2nd row: unwanted previous frame replica of a person

3rd row: Logical and operation to get rid of older replica

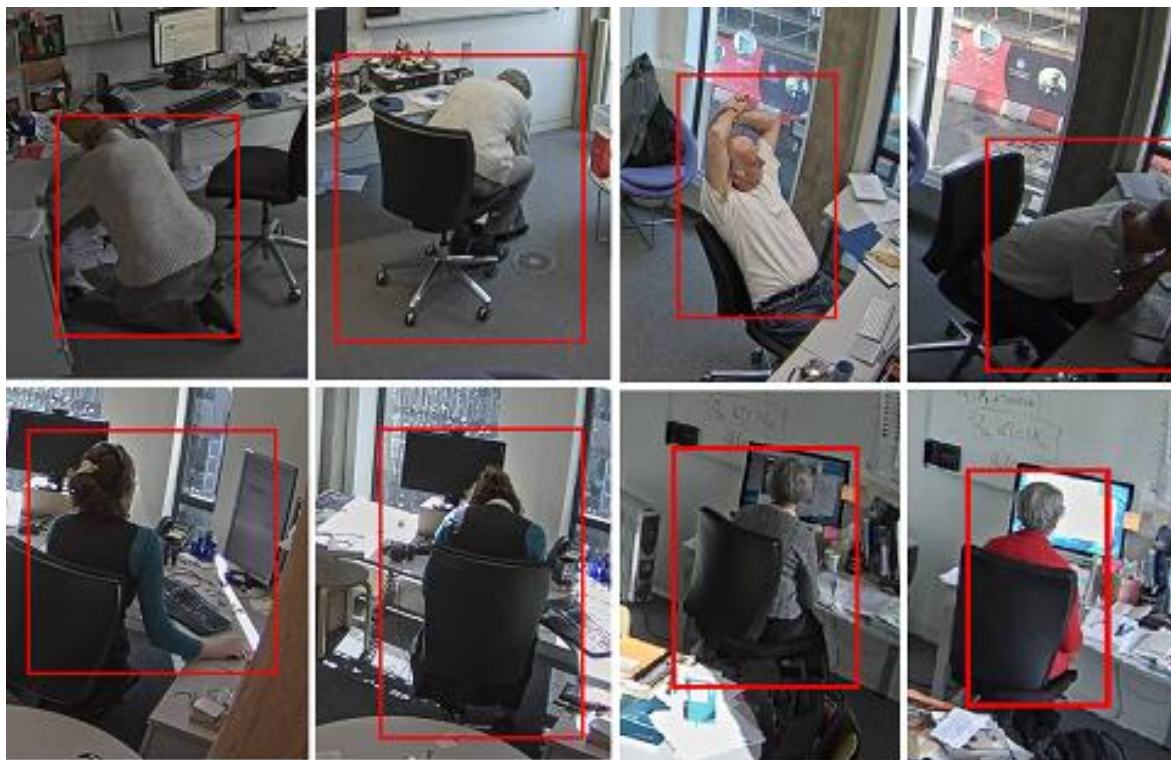
4th row: Motion cells

5th row: Bounding boxes around motion cells



Shadows and reflections problem

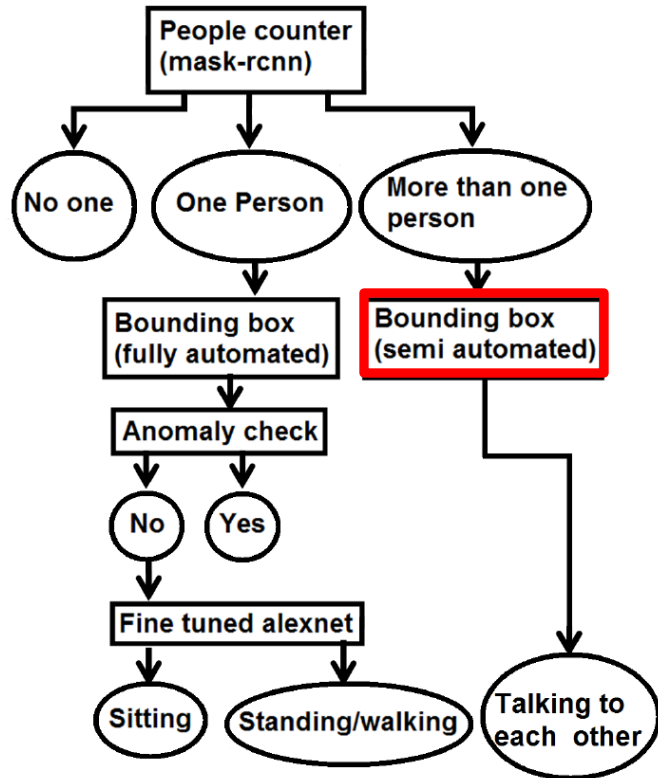
Single person bounding box method



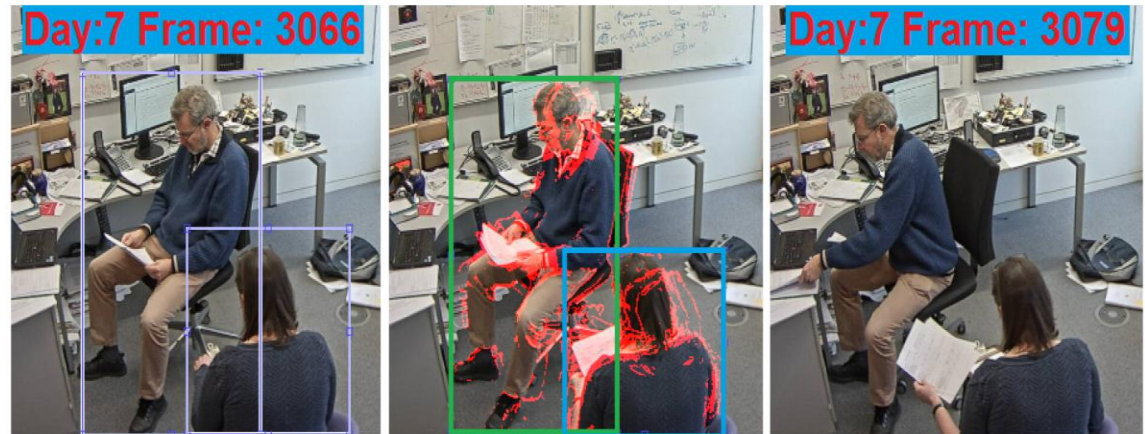
- Top row: Day 12 frame 5,651, day 12 frame 11,134, day 13 frame 15,273, day 14 frame 17,022
- Bottom row: Day 15 frame 3,971, day 15 frame 5,698, day 18 frame 1,896, day 20 frame 13659

Introduction
Overview of method
Experimental results
Conclusions

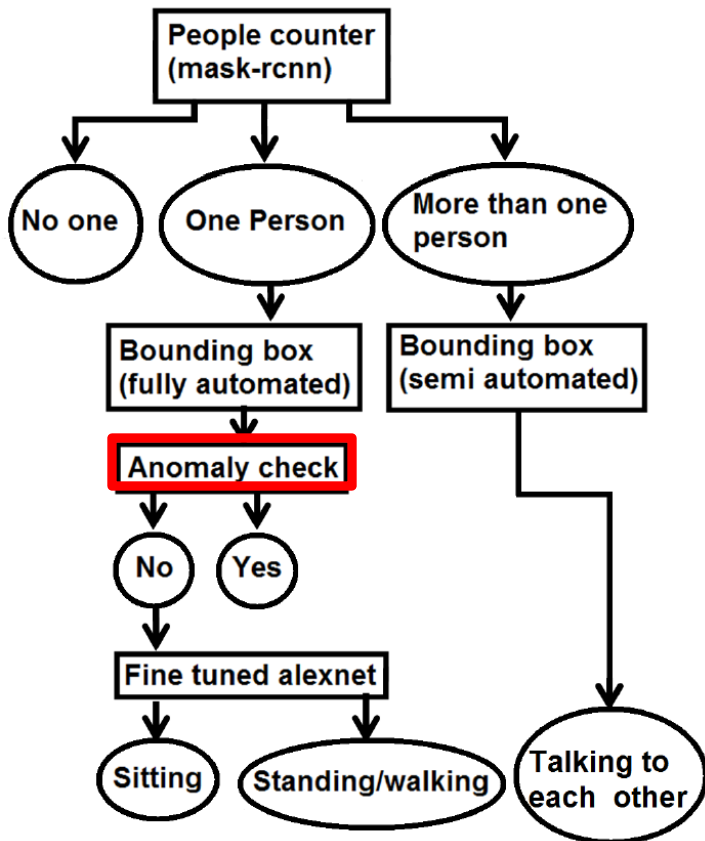
More than one person bounding box method



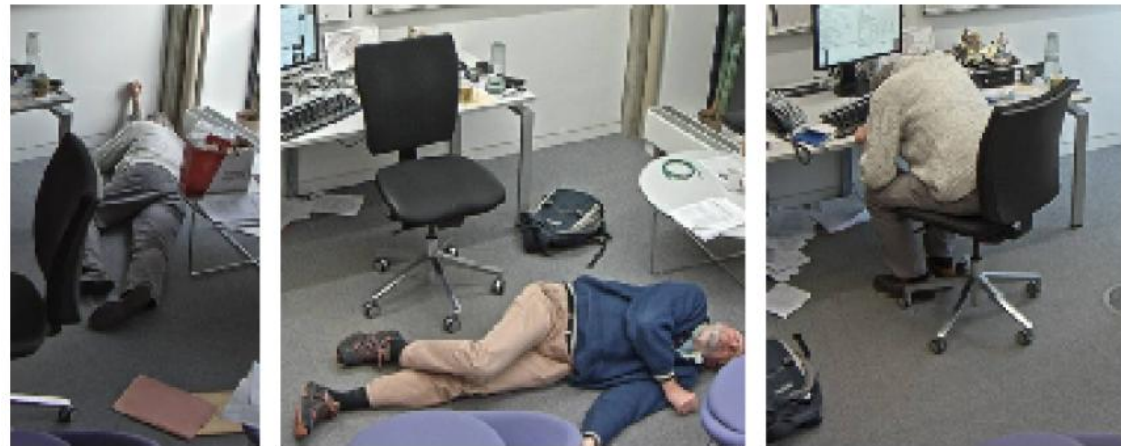
- Specialized methods for multi-person frames



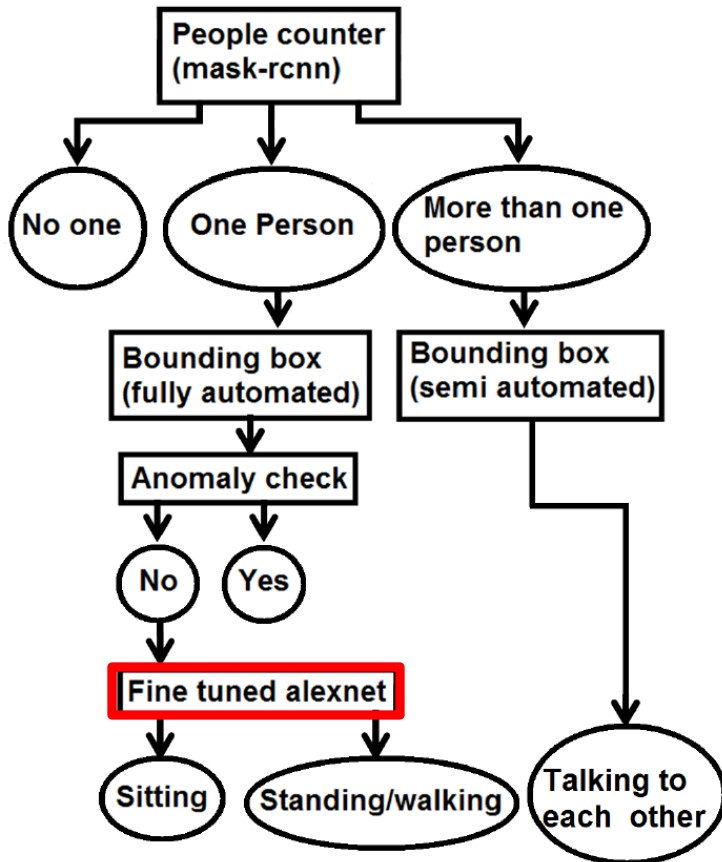
Anomaly detection (fall on floor)



- Anomalies; fall on floor & collapse on table
- Method based on detection of very low/zero cell motion as a result of the fall



Behavior labeling



- Label “*talking to each other*” assigned to frames with two or more people
- Single person frame classes: sitting and standing/walking
- A fine-tuned alexnet CNN used for classification
- Changes in labels manually verified

People counter results

Office	Manually verified true detected state changes	Manually corrected false detected state changes	Total automatic corrections	Total frames
01	274	213	7,891	236,651
02	19	15	252	54,721
03	99	41	6,251	77,628
04	87	32	8,089	87,715
Total	479	301	22,483	456,715

- Clicks required for people counting reduced to 0.71% (479 verifications and 301 corrections out of 456,715 frames)

Reduction in clicks

- Clicks for **single person bounding boxes** reduced to 0% (for 134,110 single person frames)
- Clicks for **more than one person bounding boxes** reduced to 4.02% (6,178 box initializations out of 72,650 frames requiring total 153,807 box initializations)
- Clicks for behavior labeling reduced to 0.66% (705 verifications and 953 corrections out of 249,955 single person frames)

Conclusions

- Algorithm driven approach reduces 99+% of clicks
- Combination of standard and specialized algorithms
- Applicable to low-frame rate and sparse person-watching video
- Similar reduction in click rate for 15 FPS video

Thanks