北京郵電大學
BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

BUPT-MCPRL@TRECVID 2014: Surveillance Event Detection(SED)

Qi Chen (chen_qi1990@163.com)
Zhicheng Zhao, Wenhui Jiang, Jinlong Zhao, Yuhui Huang,
Xiang Zhao, Lanbo Li, Yanyun Zhao, Fei Su, Anni Cai
BUPT-MCPRL
Beijing University of Posts and Telecommunications

Our Submission

- BUPT_MCPRL 2014 Retrospective Result

Event	Rank	ADCR	ADCR of Other Best Systems
Embrace	2	0.8318	0.8113
PeopleMeet	4	1.0354	0.8587
PeopleSplitUp	4	0.9476	0.8353
PersonRuns	4	0.9070	0.8256
Pointing	1	0.9998	1.0027

Outline

- Retrospective System Overview
- Pedestrian Detection
- Pedestrian Tracking
- Detected by CNN
  - Embrace and Pointing
- Detected by Trajectory Analysis
  - PeopleMeet and PeopleSplitUp
  - PersonRuns
- Performance Evaluation
- Conclusion

Retrospective System Overview

Pedestrian Detection by CNN

Embrace and Pointing Detection

Classified by CNN → Events Fusion

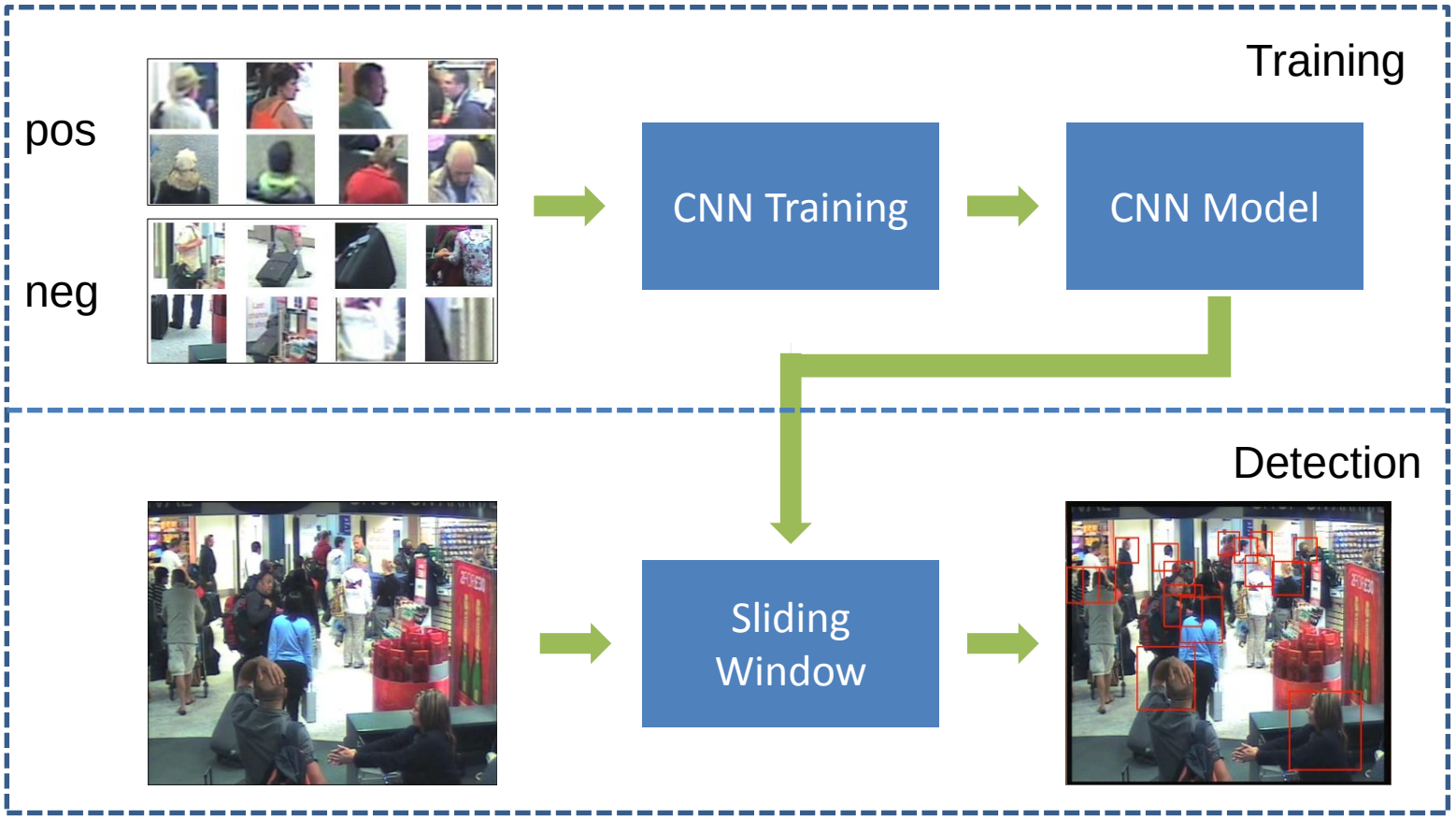
PeopleMeet, PeopleSplitUp and PersonRuns Detection

Pedestrian Tracking → Trajectory Analysis

Detections

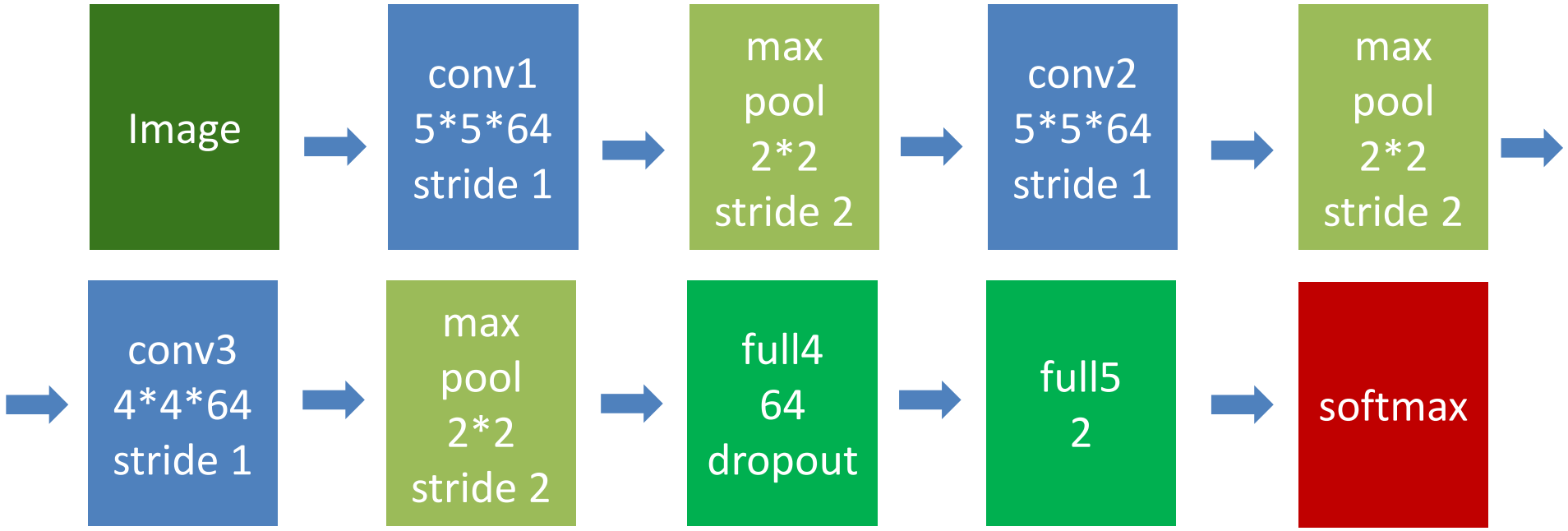
Pedestrian Detection

- Pedestrian Detection by Head-Shoulder-CNN
  - suppress the effect of partial occlusion



Pedestrian Detection

- The Architecture of Our CNN
  - much smaller than Krizhevsky's network [Krizhevsky, NIPS 2012]

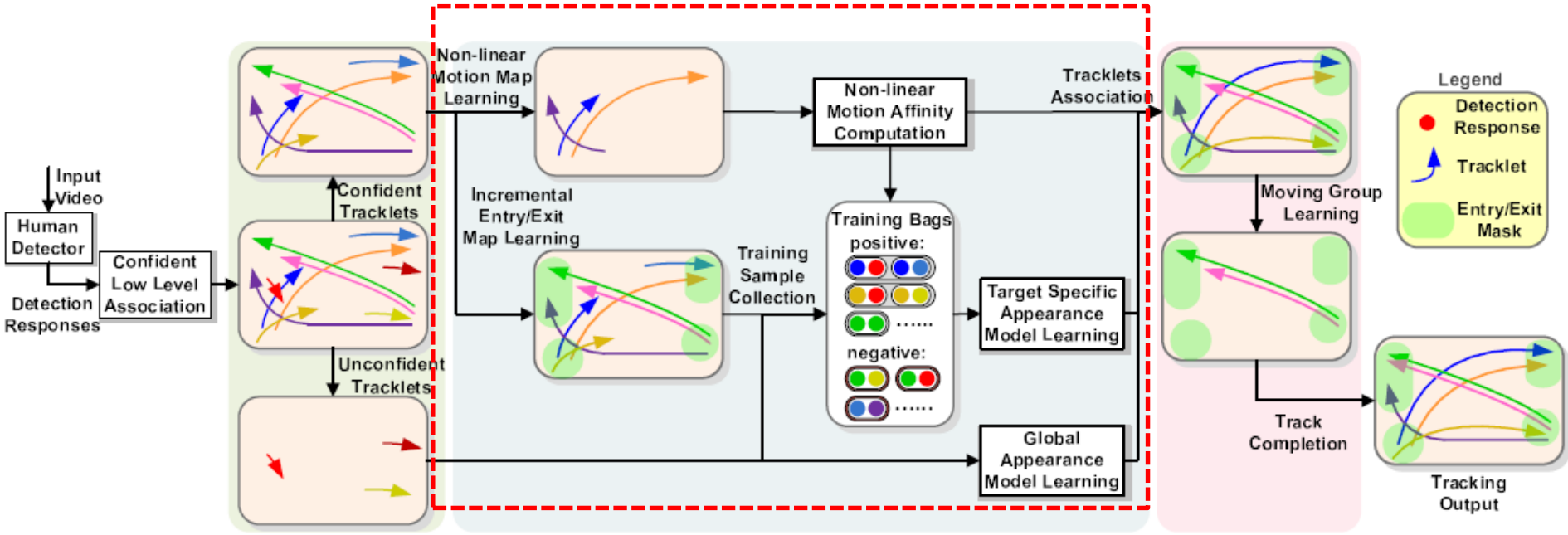


Pedestrian Detection

- Samples
  - from TrecVid08-Dev_set and TrecVid08-Eval_Set
  - positive
    - 11,538 for training
    - 4,946 for testing
    - randomly horizontal flipping
  - negative :
    - anything of non-positive
    - three times the number of positive
- Details of Training
  - single NVIDIA GTX 780Ti GPU
  - Core i7 desktop CPU
  - 3 hours for training
  - learning rate : 0.01

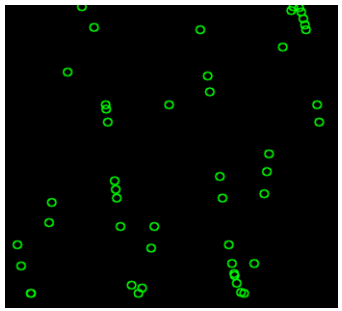
Pedestrian Tracking

- Multi-Target Tracking [Bo Yang et al. CVPR 2013]
  - online approach to learn non-linear motion patterns and robust appearance models
  - deal with detection result with long gap
  - more robust for tracking with lots of occlusion

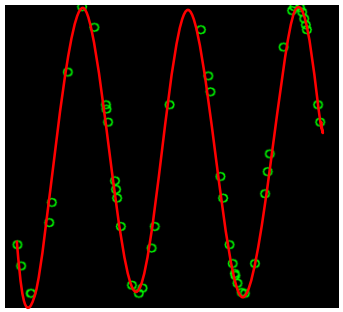


Pedestrian Tracking

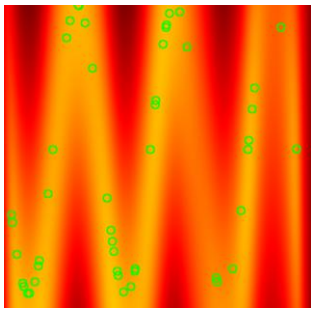
- We Propose to use Gaussian process regression to smooth the trajectory.



Detection responses x



Detection responses x and the true trajectory t



The relationship $\Pr(w|x)$ between the response x and point w of t

Unsmoothed trajectories

Smoothed trajectories

Outline

- Retrospective System Overview
- Pedestrian Detection
- Pedestrian Tracking
- Detected by CNN
  - Embrace and Pointing
- Detected by Trajectory Analysis
  - PeopleMeet and PeopleSplitUp
  - PersonRuns
- Performance Evaluation
- Conclusion

Embrace and Pointing

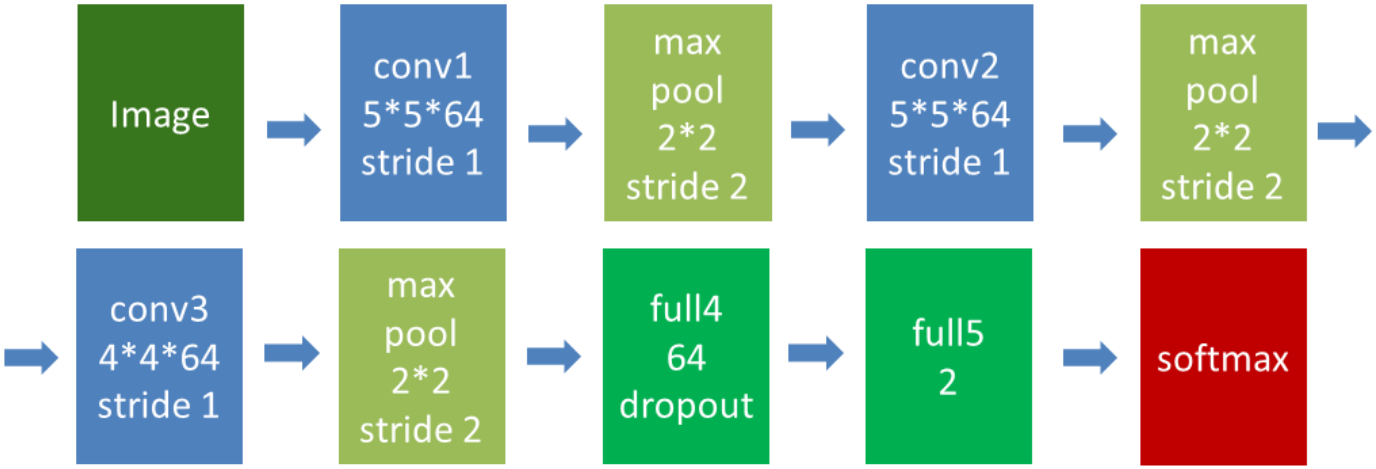
- Regard the events detection as the detection of key-poses
- Key-poses for Embrace and Pointing

Pointing

Embrace

Embrace and Pointing

- Method
  - adopt CNN to recognize the key-pose
  - use the architecture of pedestrian detection
  - the inputs of models are the pedestrian detection results with 1.5-fold expansion



The architecture of our CNN

Embrace and Pointing

- Samples
  - from TrecVid08-Dev_set and TrecVid08-Eval_Set
  - positive
    - total : 2100
    - randomly cropping
    - randomly horizontal flipping
    - RGB jittering
  - negative
    - any pedestrian detection results of non-Embrace or non-Pointing
    - three times the number of positive
- Details of Training
  - single NVIDIA GTX 780Ti GPU
  - Core i7 Desktop CPU
  - 2 hours for training
  - learning rate : 0.01

Embrace and Pointing

- retro-Embrace

Years	ADCR	MDCR	#CorDet	#FA	#Miss
2014	0.8318	0.8318	26	44	112
2013	1.0503	0.9850	13	380	162

- retro-Pointing

Years	ADCR	MDCR	#CorDet	#FA	#Miss
2014	0.9998	0.9910	21	57	774
2013	1.6387	1.0064	219	2576	844

Outline

- Retrospective System Overview
- Pedestrian Detection
- Pedestrian Tracking
- Detected by CNN
  - Embrace and Pointing
- Detected by Trajectory Analysis
  - PeopleMeet and PeopleSplitUp
  - PersonRuns
- Performance Evaluation
- Conclusion

PeopleMeet and PeopleSplitUp

- PeopleMeet
  - split into 3 subevents: walking closely, slowing down and stay
  - use HMM (Hidden Markov Model) to model the event [Chan et al. ICPR 2004]
  - observe every two persons based on their trajectories
  - the distances between persons and their speed are used as features to construct observation sequence
- PeopleSplitUp
  - split into 3 subevents : stay, speeding up, walking away
  - similar to the detection of PeopleMeet

PersonRuns

- Distinguish running trajectories
  - pick the fast-moving pedestrian tracks by Forward-backward Motion History Image (MHI) [Z Yin et al. AVPI 2009]
  - FB-MHI = F-MHI & B-MHI
  - set a threshold of the ratio of non-zero pixels in the region of the **pedestrian detection result**

Video Forward MHI Backward MHI Result

Performance Evaluation

Event	Rank	ADCR of Other Best Systems	BUPT_MCPRL 2014 Retrospective Result (Update Version)				
			ADCR	MDCR	#CorDet	#FA	#Miss
Embrace	2	0.8113	0.8318	0.8318	26	44	112
PeopleMeet	4	0.8587	1.0354	1.0018	6	128	250
PeopleSplitUp	4	0.8353	0.9476	0.9455	19	158	133
PersonRuns	4	0.8256	0.9070	0.9038	8	139	43
Pointing	1	1.0027	0.9998	0.9910	21	57	774

- Method of CNN
  - Embrace and Pointing
  - works very well
- Method of Trajectory Analysis
  - PeopleMeet, PeopleSplitUp and PersonRuns
  - not good

Conclusion

- We proposed the methods of CNN and trajectory analysis for event detection
- Method of CNN
  - works very well
  - detects a small number of false alarms and a relatively big number of correct detections
  - much less computations
  - easy to implement
- Method of trajectory analysis
  - not good
  - difficult to get the true information such as velocity

Thanks!

www.bupt-mcprl.net