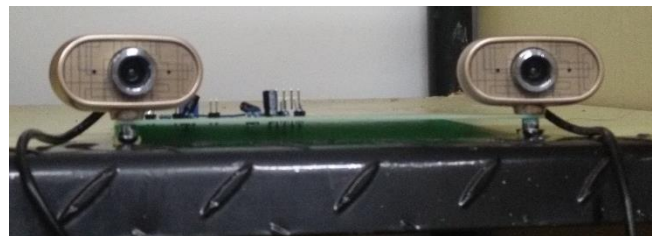
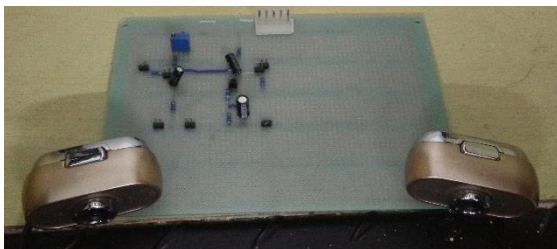





基于双目系统的目标跟踪与预测

行人检测

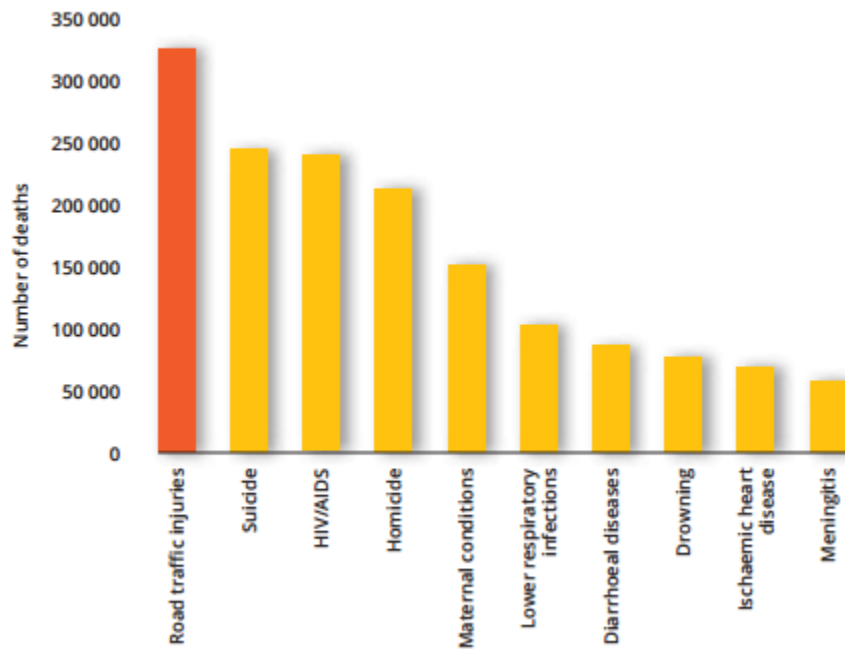


● 2015全球道路安全报告

● 1  () / week \approx 1  / week

● 2  () >  ()

Top ten causes of death among people aged 15–29 years, 2012



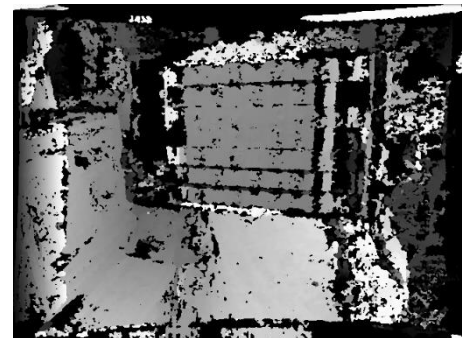
http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/

大纲

- 背景
- 计划+分工:
 - Particle Filter
 - Stixel World, HOG+SVM
 - 双目信息的效果
 - faster-rcnn
 - Rich feature 效果
- 总结遇到的问题
- 成果
- 规划
 - 双目/尺度/遮挡
- 参考文献

分工

- 增辉:
 - Particle Filter
 - OpenCV
 - Computer Vision(cs4495)
- 启睿
 - Stereo Camera
 - Machine Learning(cs229)
 - Stixel World, HOG+SVM
- 兴路
 - Faster-rcnn
 - ConvNet(cs231n)



```
video_file = os.path.join(cfg.DATA_DIR, 'demo', video_name)
# in = cv2.imread(video_file)
cap=cv2.VideoCapture(video_file)
ret, frame = cap.read()
if not ret:
    raise KeyError
in_c = domnet(frame)
while in_c is None:
    ret, frame = cap.read()
    if not ret:
        raise KeyError
    in_c = domnet(frame)
height, width, layers = in_c.shape
out = cv2.VideoWriter('demo.avi', cv2.cv.CV_FOURCC('X', 'V', 'I', 'D'), 40, (width, height))
# global ret
while(True):
    ret, frame=cap.read()
    if not ret:
        break
```

```
4500 Detection took 0.307s for 300 object proposals
4501 Detection took 0.319s for 300 object proposals
4502 Detection took 0.318s for 300 object proposals
4503 Detection took 0.300s for 300 object proposals
4504 Detection took 0.323s for 300 object proposals
4505 Detection took 0.314s for 300 object proposals
4506 Detection took 0.303s for 300 object proposals
4507 Detection took 0.322s for 300 object proposals
4508 Detection took 0.315s for 300 object proposals
4509 Detection took 0.322s for 300 object proposals
4510 Detection took 0.304s for 300 object proposals
4511 Detection took 0.322s for 300 object proposals
4512 Detection took 0.312s for 300 object proposals
4513 Detection took 0.302s for 300 object proposals
4514 Detection took 0.327s for 300 object proposals
4515 Detection took 0.315s for 300 object proposals
4516 Detection took 0.308s for 300 object proposals
4517 Detection took 0.324s for 300 object proposals
4518 Detection took 0.310s for 300 object proposals
4519 Detection took 0.308s for 300 object proposals
4520
```

Particle Filter

以追踪人脸patch为例



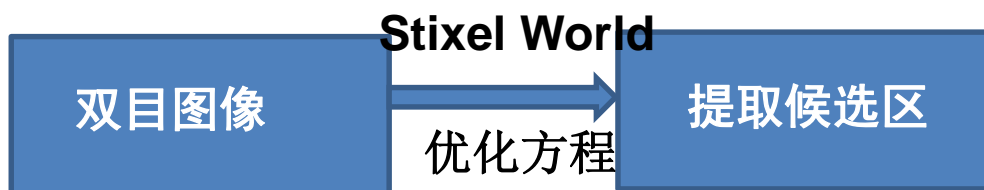
Stixel World, Idea

- 如何利用双目信息？
 - 一般思路：



Stixel World, Idea

- 利用双目优点：
 - 适用于室外场景



- 避开缺点：
 - 匹配速度慢
 - 视差图质量差

Stixel World, Objectness Estimation

原理：是行人+脚下是地面
(详见Supplementary Material)

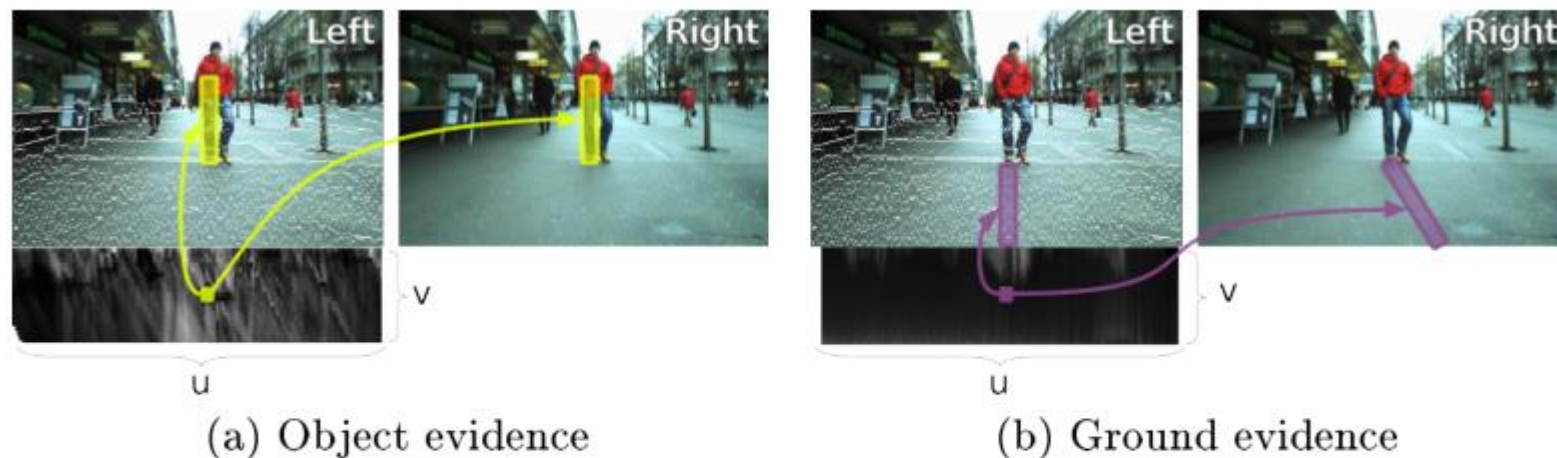


Figure 3: The object and ground costs are computed by matching pixels in the left and right images. White dots on the image indicates object-ground boundary candidates, based on horizontal gradient maxima.

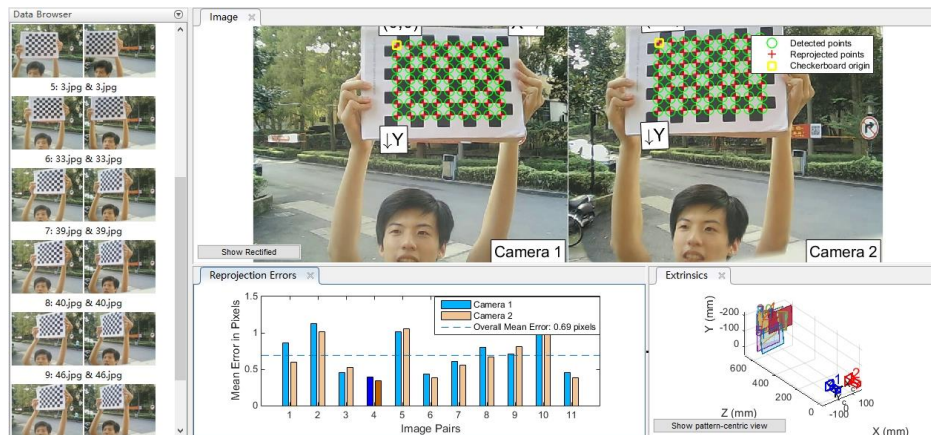
Stixel World, Hog + SVM

效果:

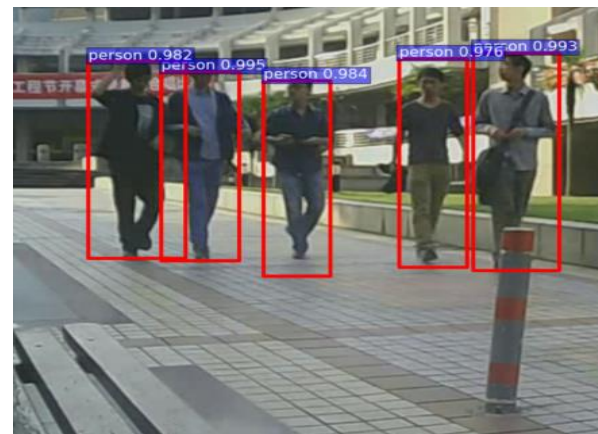


利用先验

- 先验=人和地面的位置关系+人的身高

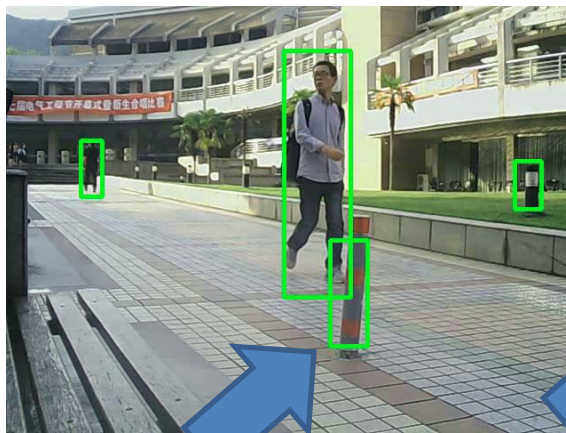


利用先验



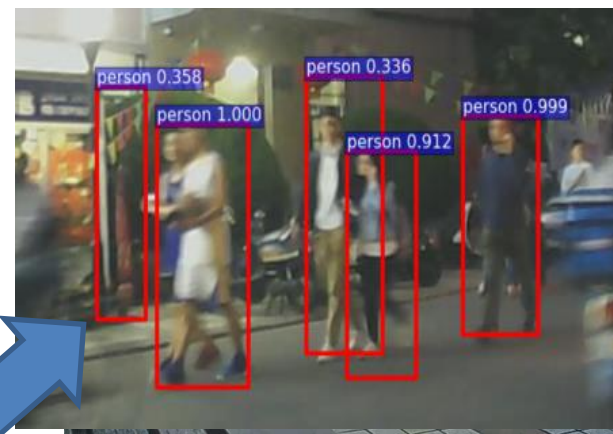
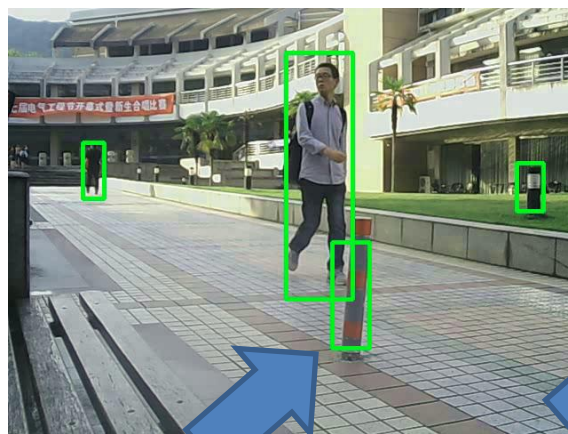
Stixel World, Hog + SVM

问题出在哪里？



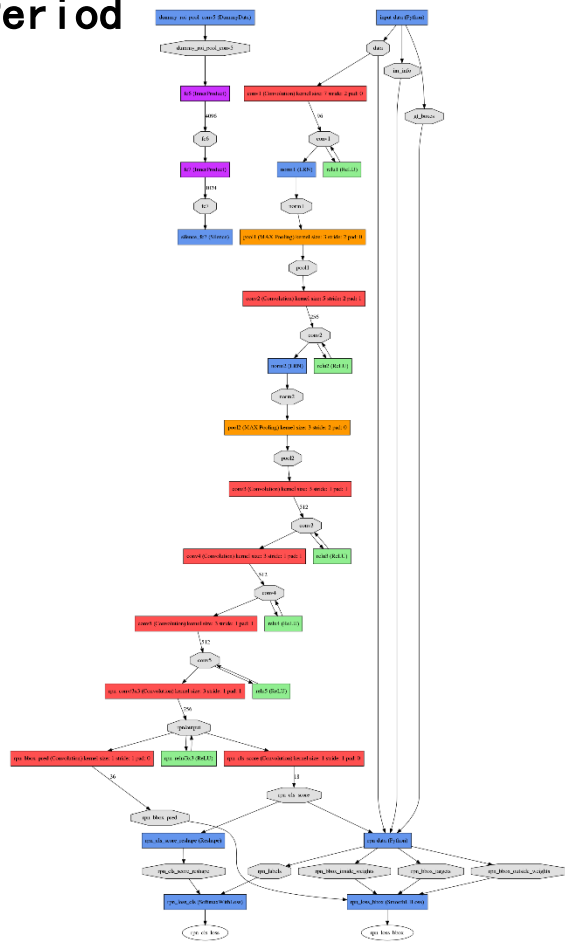
Stixel World, Hog + SVM

Hog特征 饱和

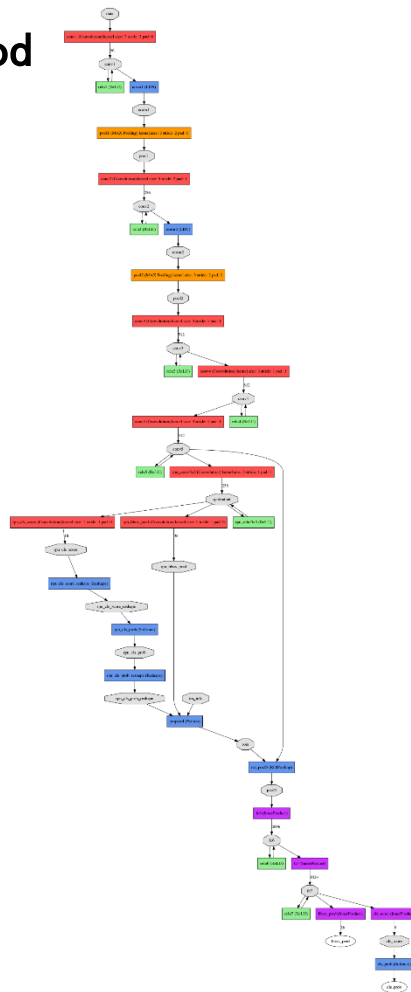


Faster-rcnn

Train Period



Test Period



详见Supplementary Material

选取数据集

- KITTI's page(car,pedestrian, etc.):
http://www.cvlibs.net/datasets/kitti/eval_object.php/

- 数据量大
- 场景单一



- Caltech-USA:
http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

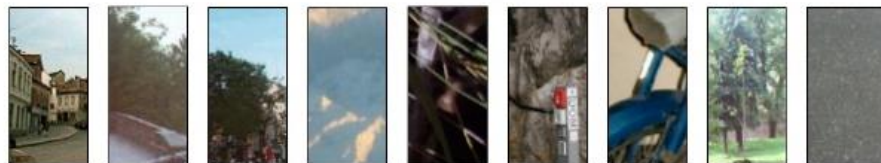
- 场景单一



制作数据集



Positive sample



Negative sample



ETHZ1

ETHZ2

ETHZ3

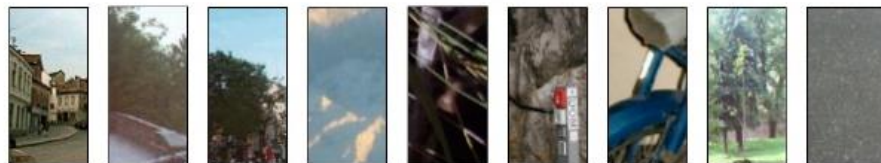
- INRIA: <http://pascal.inrialpes.fr/data/human/>
 - 数据量不足
 - 场景丰富

```
charm. x screen-r drone luzai@luzai-PC:~/luzai/faster-rcnn-drone
nexus_video5_lng41.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng41.png
nexus_video5_lng42.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng42.png
nexus_video5_lng43.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng43.png
nexus_video5_lng44.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng44.png
nexus_video5_lng45.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng45.png
nexus_video5_lng46.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng46.png
nexus_video5_lng47.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng47.png
nexus_video5_lng48.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng48.png
nexus_video5_lng49.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng49.png
nexus_video5_lng4.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng4.png
nexus_video5_lng50.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng50.png
nexus_video5_lng51.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng51.png
nexus_video5_lng52.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng52.png
nexus_video5_lng53.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng53.png
nexus_video5_lng54.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng54.png
nexus_video5_lng55.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng55.png
nexus_video5_lng56.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng56.png
nexus_video5_lng57.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng57.png
nexus_video5_lng58.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng58.png
nexus_video5_lng59.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng59.png
nexus_video5_lng60.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng60.png
nexus_video5_lng61.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng61.png
nexus_video5_lng62.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng62.png
nexus_video5_lng63.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng63.png
nexus_video5_lng64.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng64.png
nexus_video5_lng65.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng65.png
nexus_video5_lng66.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng66.png
nexus_video5_lng67.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng67.png
nexus_video5_lng68.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng68.png
nexus_video5_lng69.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng69.png
nexus_video5_lng70.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng70.png
nexus_video5_lng7.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng7.png
nexus_video5_lng8.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng8.png
nexus_video5_lng9.png -> /home/luzai/luzai/faster-rcnn-drone/output/videos/nexus/video5/lng9.png
ImageSets
3 directories, 338 files
-> stanford_campus_dataset.ln ls
Annotations Images ImageSets
-> stanford_campus_dataset.ln
```


制作数据集



Positive sample



Negative sample



```
# print 'Loading: {}'.format(filename)
```

```
with open(filename) as f: f: <closed file '/home/luzai/luzai/py-faster-rcnn/data/INRIA_Person_devkit/data/Annotati
    data = f.read() data: '# PASCAL Annotation Version 1.00\n\nImage filename : "Train/pos/crop_000010.png"\nImage
import re 0: '(194, 127) - (413, 647)' __len__: 1 re: <module 're' from '/usr/lib/python2.7/re.pyc'>
objs = re.findall('\\(\\d+, \\d+\\)[\\s-]+\\(\\d+, \\d+\\)', data) objs: <type 'list': ['(194, 127) - (413, 647)']
```

```
num_objs = len(objs) num_objs: 1
```

```
boxes = np.zeros((num_objs, 4), dtype=np.uint16) boxes: [[194 127 413 647]]
```

```
gt_classes = np.zeros((num_objs), dtype=np.int32) gt_classes: [1]
```

```
overlaps = np.zeros((num_objs, self.num_classes), dtype=np.float32) overlaps: [[ 0.  1.]]
```

```
# "Seg" area here is just the box area
```

```
seg_areas = np.zeros((num_objs), dtype=np.float32) seg_areas: [ 114620.]
```

```
# Load object bounding boxes into a data frame.
```

```
for ix, obj in enumerate(objs): ix: 0 obj: '(194, 127) - (413, 647)'
```

```
# Make pixel indexes 0-based
```

```
coor = re.findall('\\d+', obj) coor: <type 'list': ['194', '127', '413', '647']
```

```
x1 = float(coor[0]) x1: 194.0
```

```
y1 = float(coor[1]) y1: 127.0
```


Finetune

ImageNet训练



From Caffe Model Zoo

- Inria Finetune
 - RPN
 - FC layer

```
# print 'Loading: {}'.format(filename)
with open(filename) as f: f: <closed file '/home/luzai/luzai/py-faster-rcnn/data/INRIA_Person_devkit/data/Annotati
    data = f.read() data: '# PASCAL Annotation Version 1.00\n\nImage filename : "Train/pos/crop_000010.png"\nImage
import re 0: '(194, 127) - (413, 647)' __len__: 1 re: <module 're' from '/usr/lib/python2.7/re.pyc'>
objs = re.findall('\\(\\d+, \\d+\\)[\\s-]+\\(\\d+, \\d+\\)', data) objs: <type 'list': ['(194, 127) - (413, 647)']

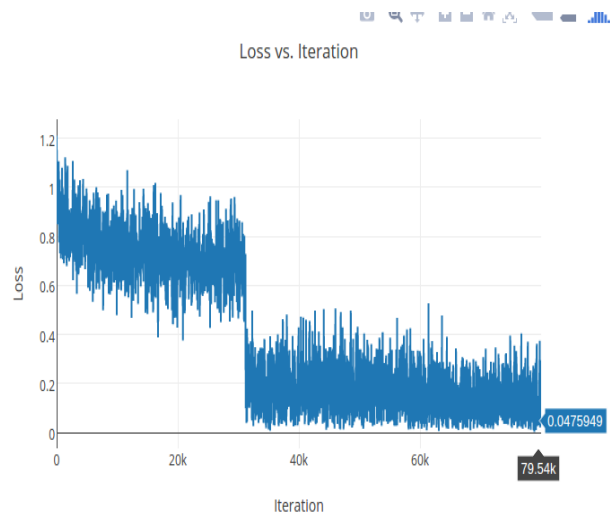
num_objs = len(objs) num_objs: 1

boxes = np.zeros((num_objs, 4), dtype=np.uint16) boxes: [[194 127 413 647]]
gt_classes = np.zeros((num_objs), dtype=np.int32) gt_classes: [1]
overlaps = np.zeros((num_objs, self.num_classes), dtype=np.float32) overlaps: [[ 0.  1.]]

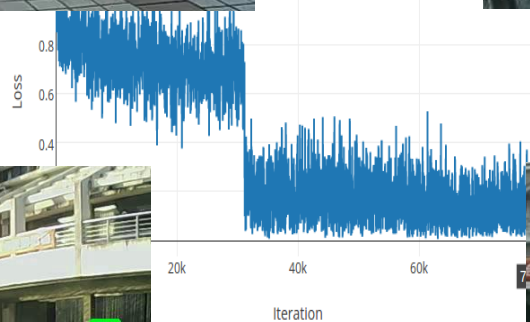
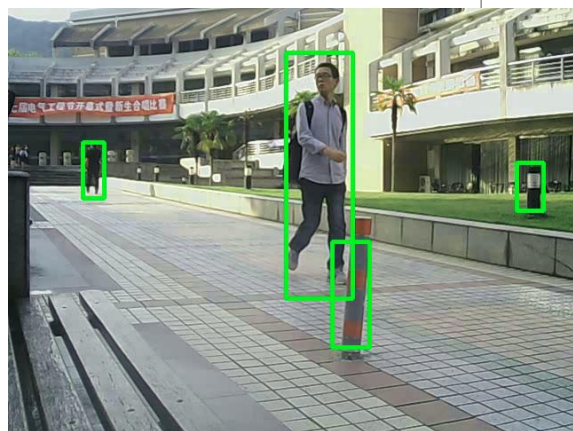
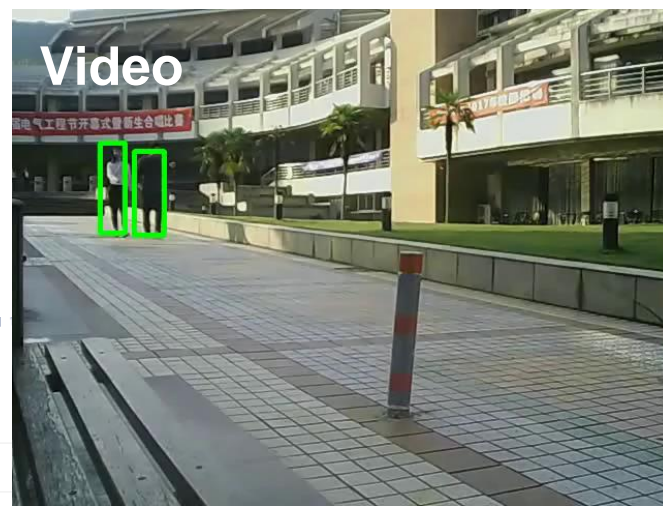
# "Seg" area here is just the box area
seg_areas = np.zeros((num_objs), dtype=np.float32) seg_areas: [ 114620.]

# Load object bounding boxes into a data frame.
for ix, obj in enumerate(objs): ix: 0 obj: '(194, 127) - (413, 647)'
    # Make pixel indexes 0-based
    coor = re.findall('\\d+', obj) coor: <type 'list': ['194', '127', '413', '647']
    x1 = float(coor[0]) x1: 194.0
    y1 = float(coor[1]) y1: 127.0
```

交替训练结果



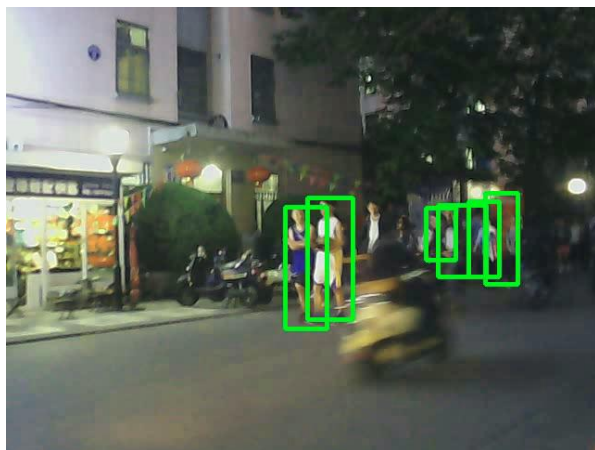
改进效果



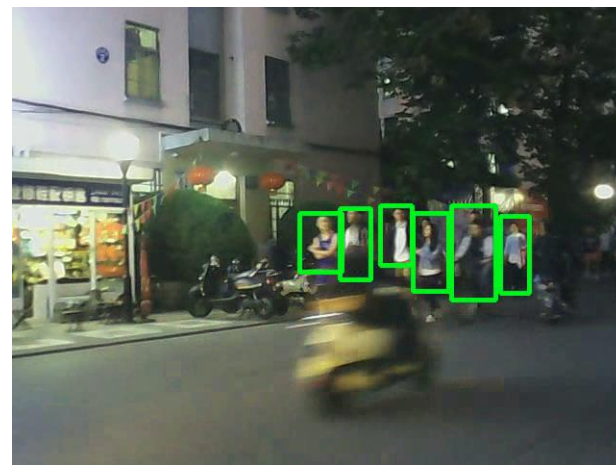
大纲

- 背景
- 计划+分工:
 - Particle Filter
 - Stixel World, HOG+SVM
 - 双目信息的效果
 - faster-rcnn
 - Rich feature 效果
- 总结遇到的问题
- 成果
- 规划
 - 双目/尺度/遮挡
- 参考文献

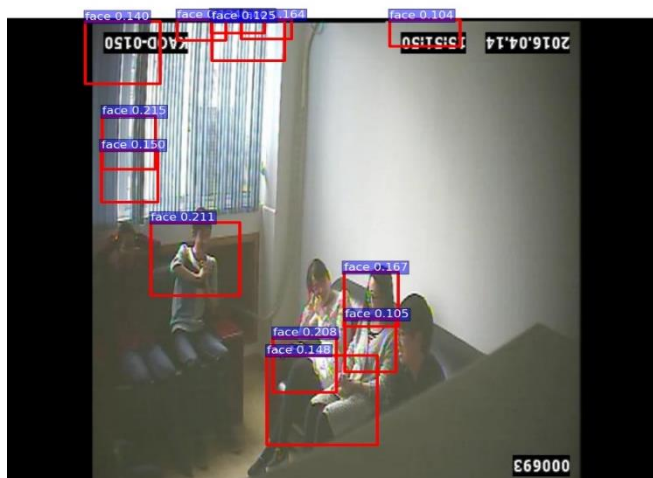
仍存在问题



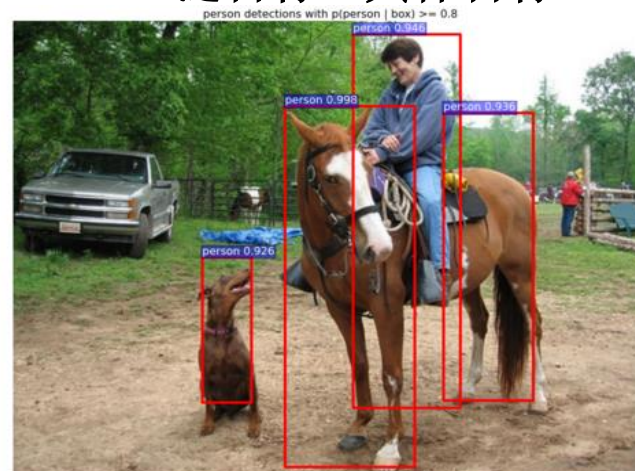
时序建模



应用于实际监控场景



泛目标->具体目标



规划

- 已有模型

- Stixel World

- 简洁的模型
 - 合理利用深度信息

- Faster-rcnn

- 泛目标检测框架
 - 更好的特征提取

- 理论上突破

- 深度信息

- 用于提取候选区
 - 建立自己的优化方程
 - 设计Rgb-d输入的RCNN
 - 视差信息输入
 - 能否End-to-End?
 - 干预CNN学习过程
 - 设计Loss 函数
 - 可视化CNN, 观察(抑制)一些激活值

- 时序建模

- GRU-RCN(LSTM)

规划

- 已有模型
 - Stixel World
 - 简洁的模型
 - 合理利用深度信息
 - Faster-rcnn
 - 泛目标检测框架
 - 更好的特征提取
- 应用上突破
 - 行人再识别
 - Siamese Instant Search 思路
 - 模型压缩
 - 人脸超分辨重建
 - 行人性格分析

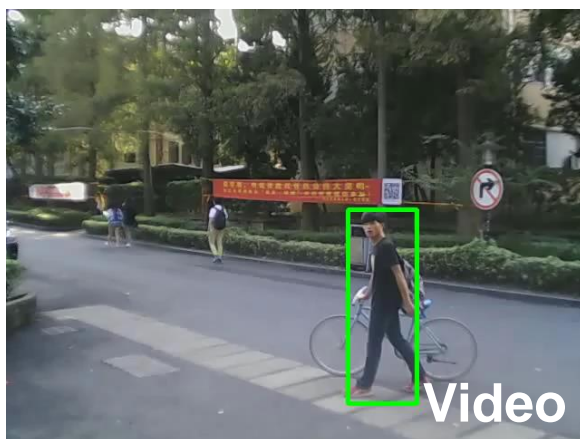
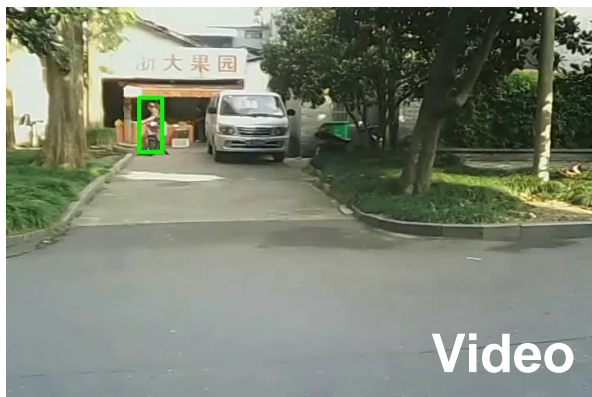
Reference

- **Benenson, R., et al., Pedestrian detection at 100 frames per second. 2012, IEEE. p. 2903 - 2910.**
- Benenson, R., R. Timofte and L. Van Gool, Stixels estimation without depth map computation, in 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). 2011. p. 2010 - 2017.
- Opitz, M., et al., Grid Loss: Detecting Occluded Faces. 2016
- Tao, R., E. Gavves and A.W.M. Smeulders, Siamese Instance Search for Tracking. 2016
- Reed, S., et al., Learning Deep Representations of Fine-grained Visual Descriptions. 2016.
- INRIA: <http://pascal.inrialpes.fr/data/human/>
- KITTI's page(car,pedestrian, etc.):
http://www.cvlibs.net/datasets/kitti/eval_object.php/
- <https://bitbucket.org/rodrigob/doppia/overview>
- <http://vision.ucsd.edu/~pdollar/index.html>
- https://github.com/ShaoqingRen/faster_rcnn/

Reference

- He, K., et al., Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans Pattern Anal Mach Intell, 2015. 37(9): p. 1904-16.
- **Ren, S., et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell, 2016.**
- Xu, J., et al., 2D/3D Sensor Exploitation and Fusion for Enhanced Object Detection. 2014, IEEE. p. 778 - 784.
- Im, D.J., et al., Generating images with recurrent adversarial networks. 2016.
- Reed, S., et al., Generative Adversarial Text to Image Synthesis. 2016.
- Mahendran, A. and A. Vedaldi, Understanding Deep Image Representations by Inverting Them. 2014.
- Denton, E., et al., Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. 2015.
- Larsen, A.B.L., et al., Autoencoding beyond pixels using a learned similarity metric. 2015.

成果



Thank You!