视觉

机器学习与人工智能（三）

陈一帅

yschen@bjtu.edu.cn

北京交通大学电子信息工程学院

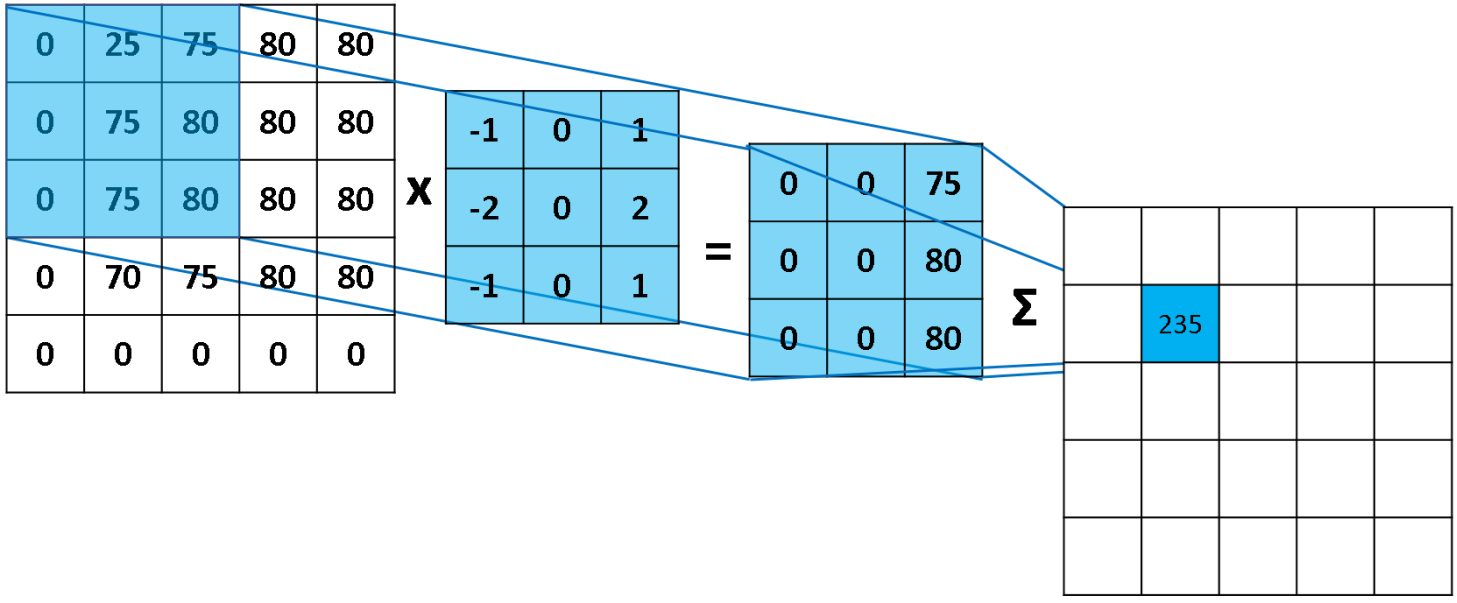
内容

- 卷积与滤波
- 经典方法
- 深度学习方法
- 目标检测和识别
- 问题

卷积和滤波

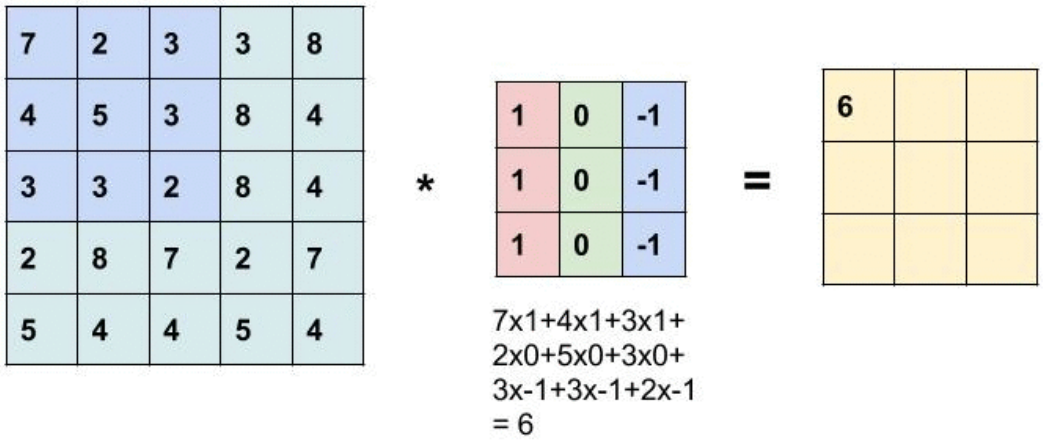
卷积和滤波

- 二维卷积
- 将卷积核与图形对应位置的像素值相乘，然后相加



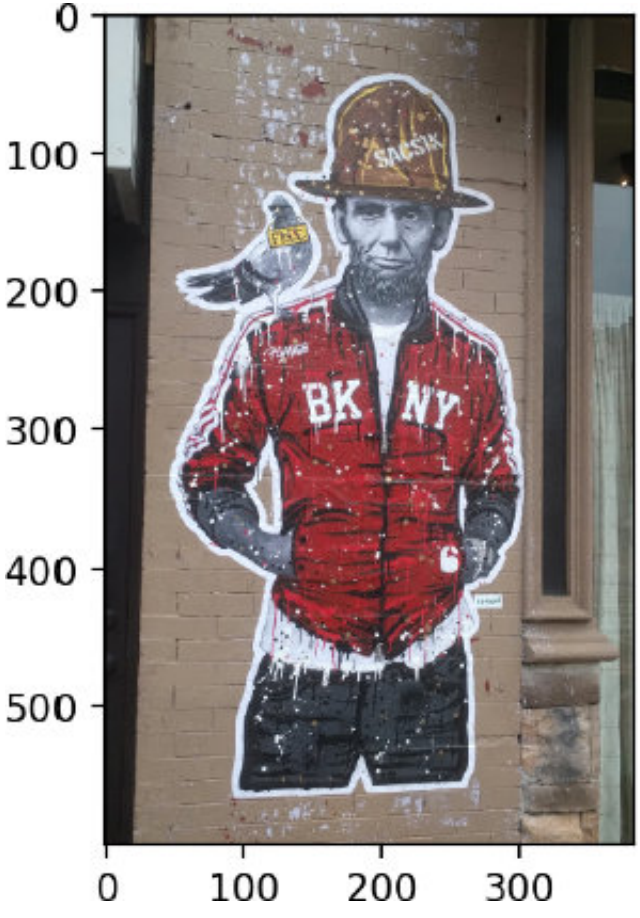
图像卷积

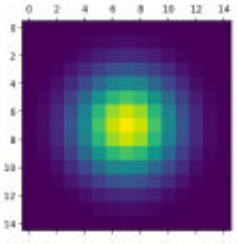
- 卷积核在图片上滑动，进行卷积操作

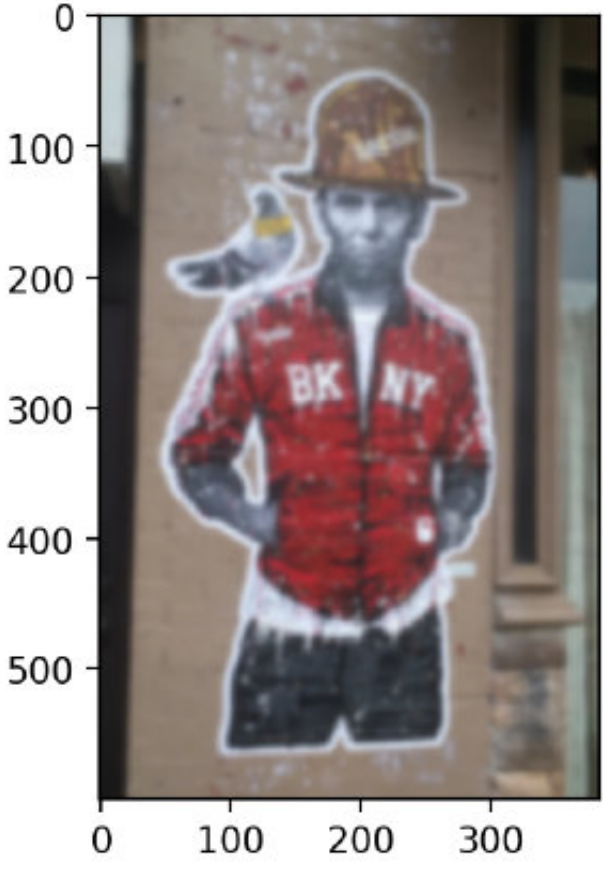


卷积实现图像平滑

模糊化

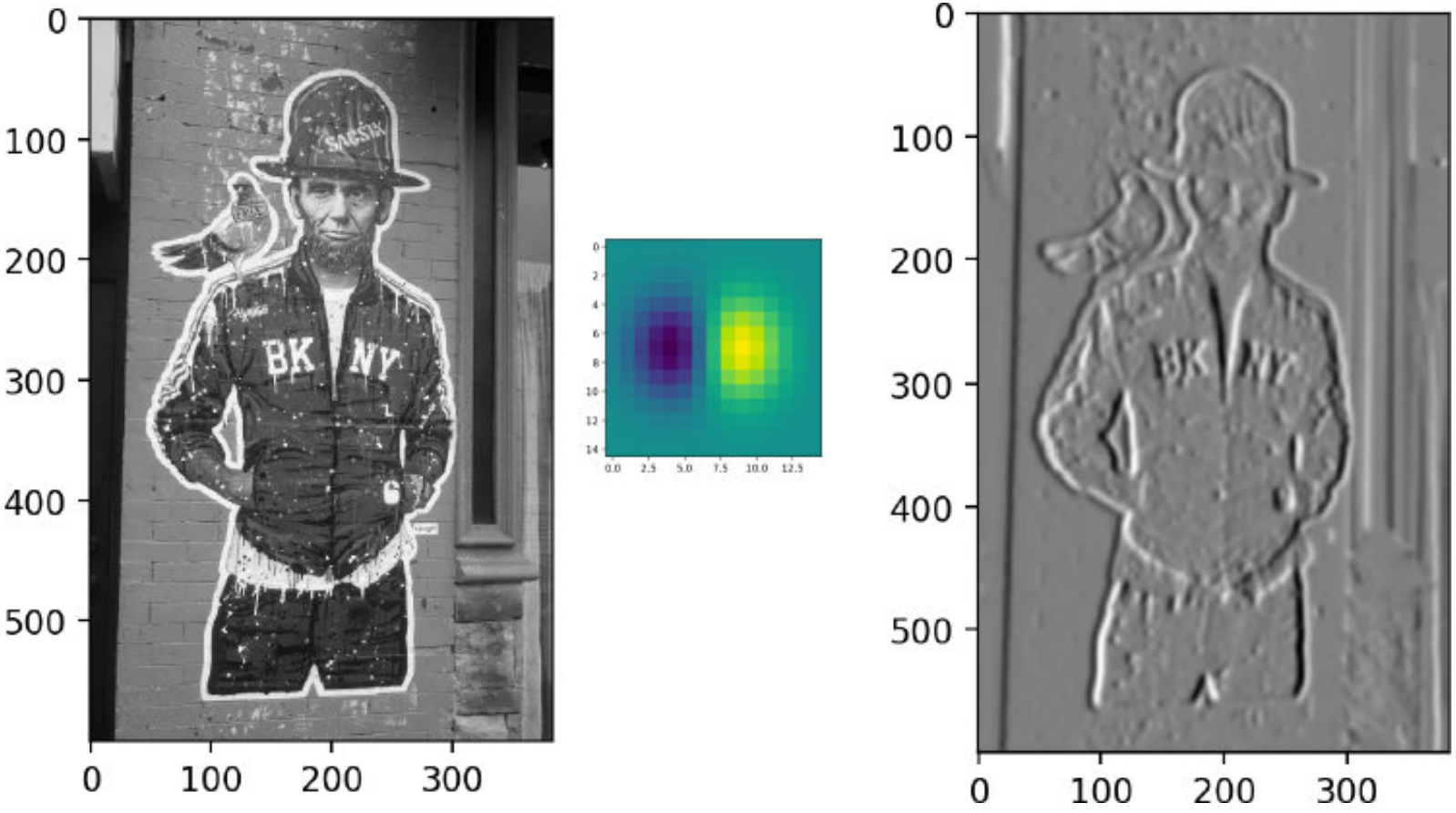






卷积获得图像梯度

提取边缘



经典方法

Feature Extraction

HOG、SIFT、Surf

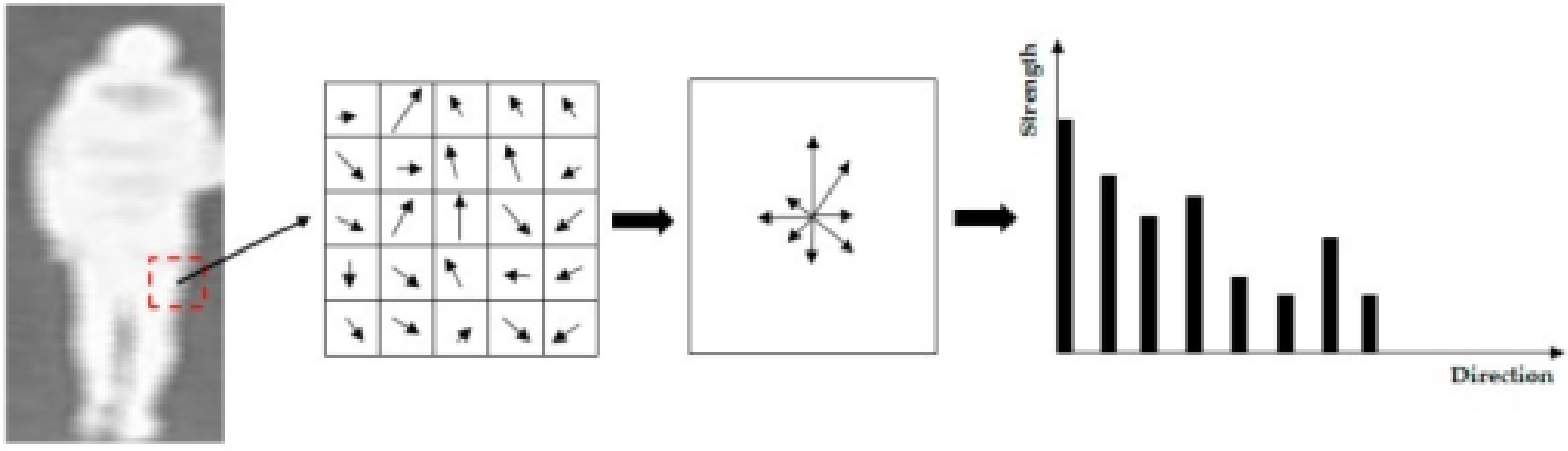
图像HOG特征

方向梯度直方图

Histogram of oriented gradient

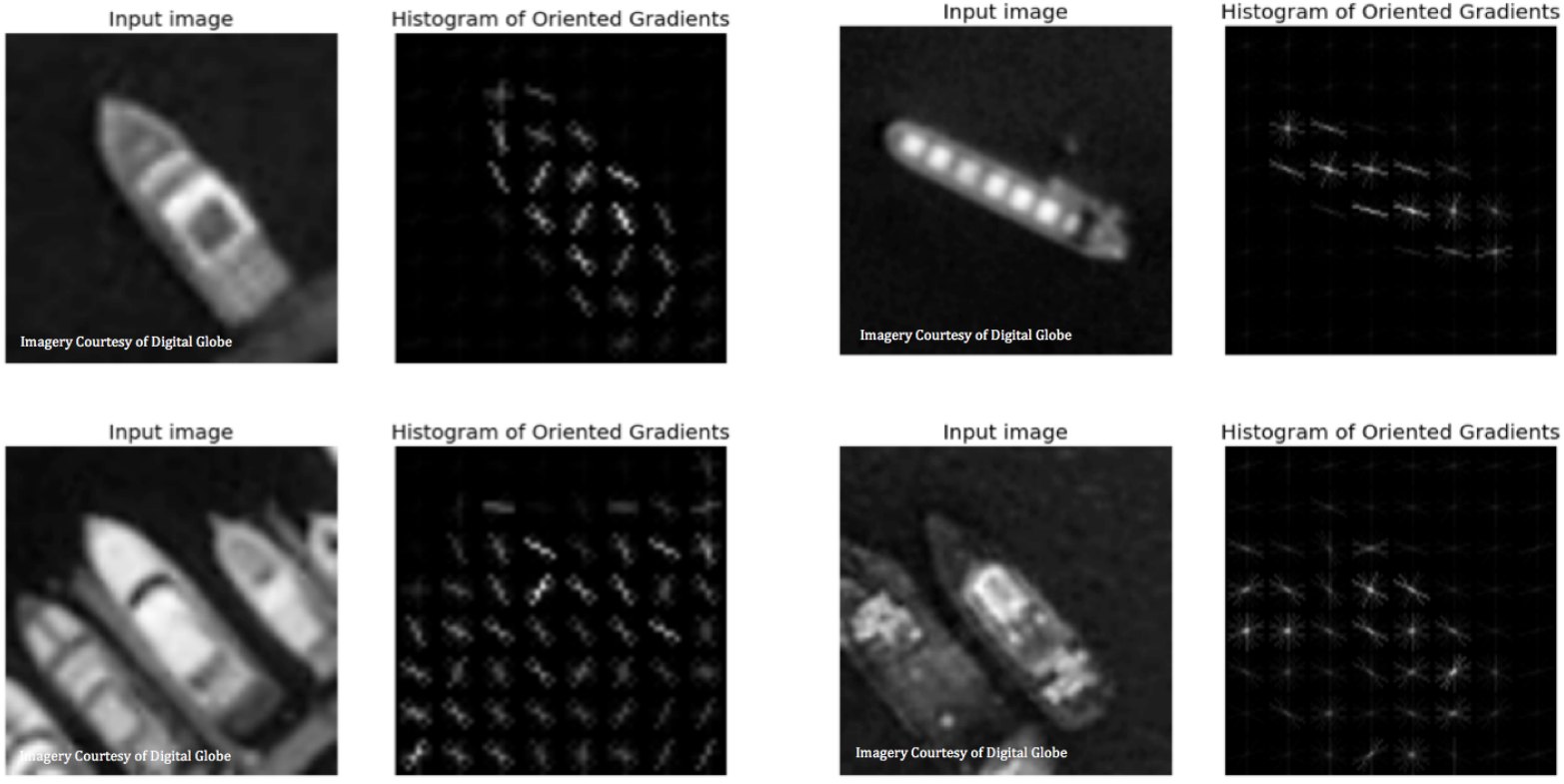
图像HOG特征

- 2005年Navneet Dalal和Bill Triggs，CVPR发表
- 适合做行人检测
- 行人直立，细微肢体动作不影响检测效果



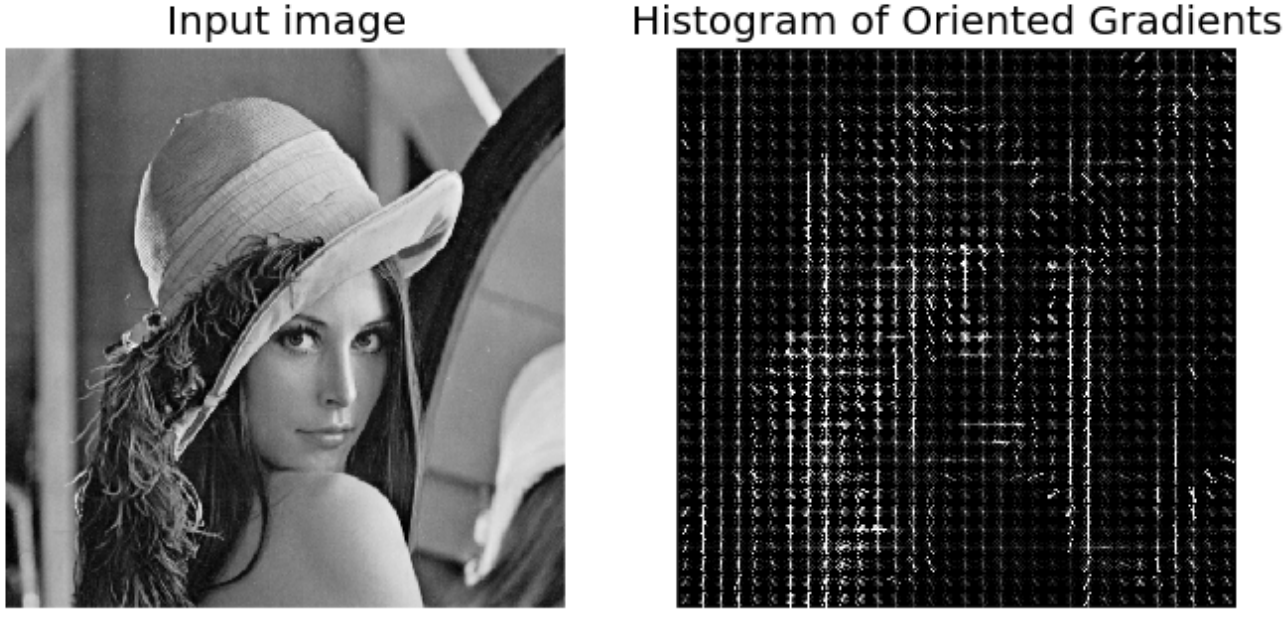
HOG效果

描述图像中目标的外表和形状



实现

1. 将图像分成小区，计算区中像素梯度或边缘方向
2. 统计直方图，构成特征描述
3. 归一化，应对光照变化和阴影



SIFT算法

关键点检测、描述

SIFT

- 尺度不变特征转换
- Scale-invariant feature transform
- 广泛应用于目标识别
  - 3个以上 SIFT特征就足以计算出目标的位置与方位
- David Lowe1999年发表，2004年完善总结

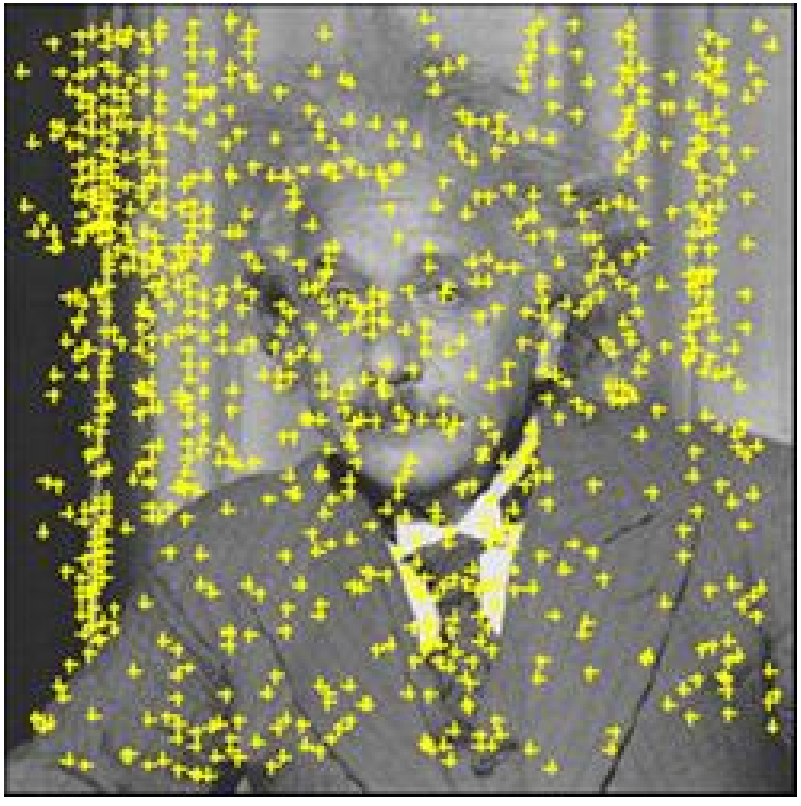
SIFT思想

- 找到关键点的位置、尺寸、方向

1）关键点检测

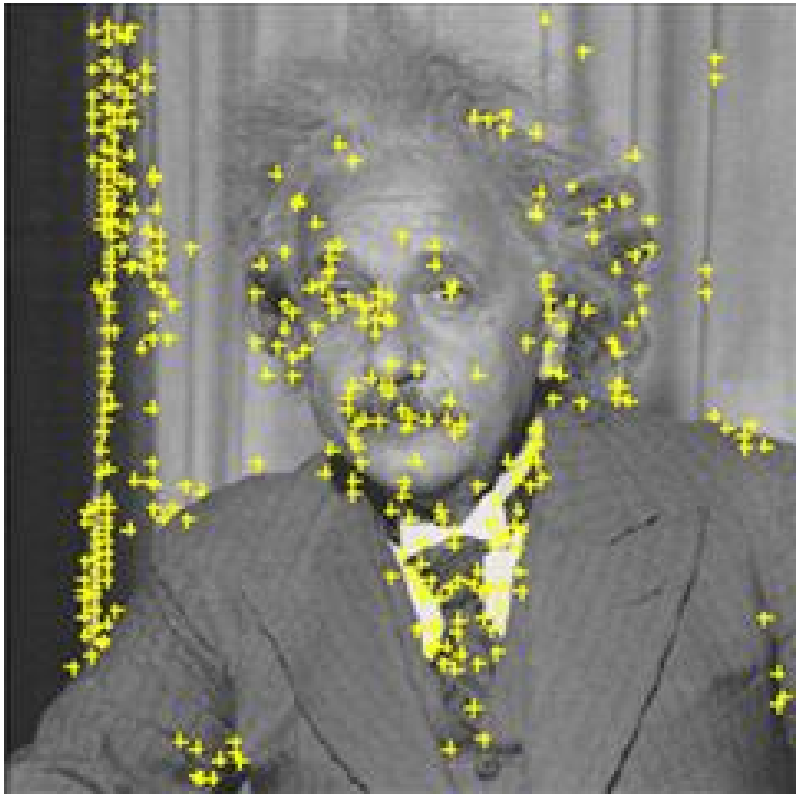
a）尺度空间的极值侦测

- 图像在不同尺度下用高斯滤波进行卷积
- 利用卷积结果的差异找出关键点



b）关键点定位

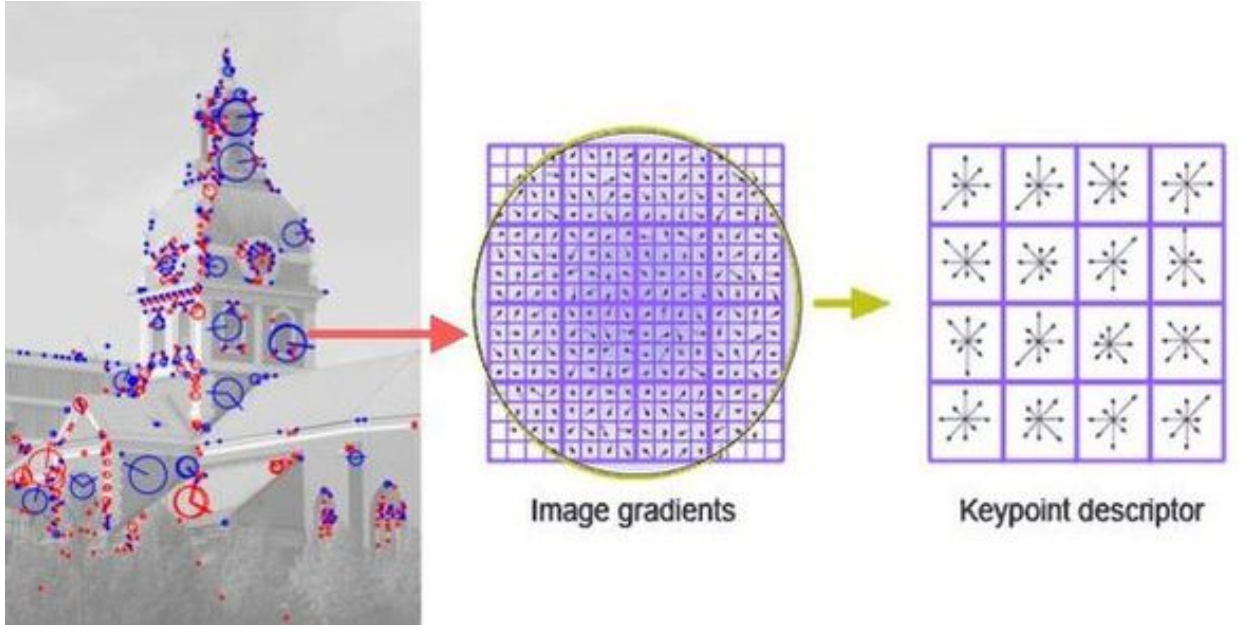
- 由关键点附近像素信息、关键点尺寸、主曲率，筛选关键点
- 消除边缘或易受噪声干扰的关键点



2）关键点描述

a）方位定向

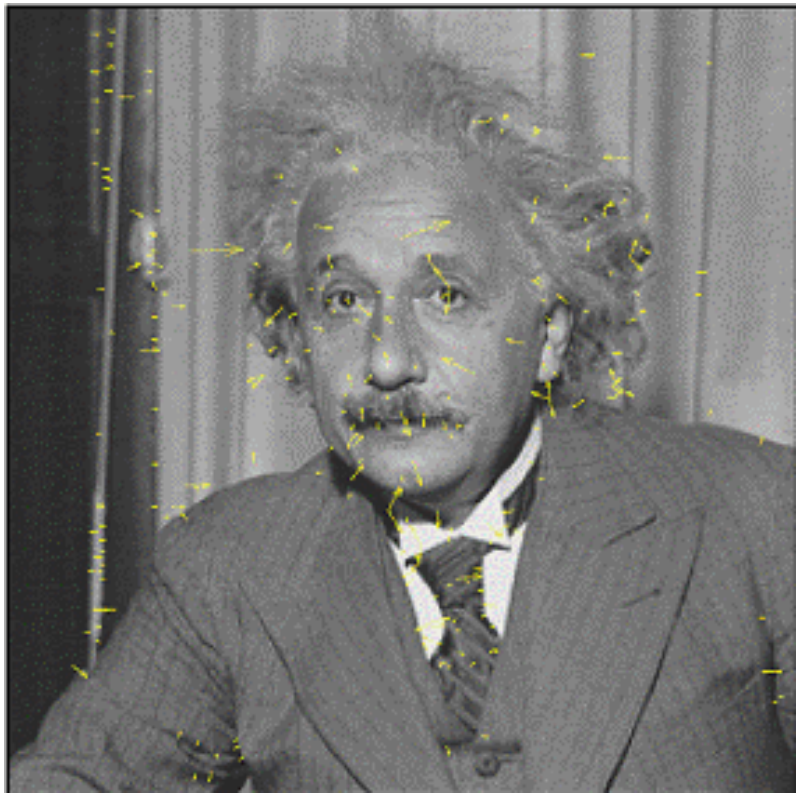
- 以关键点相邻像素的梯度方向分布作为方向参数
- 使关键点描述子具备旋转不变性



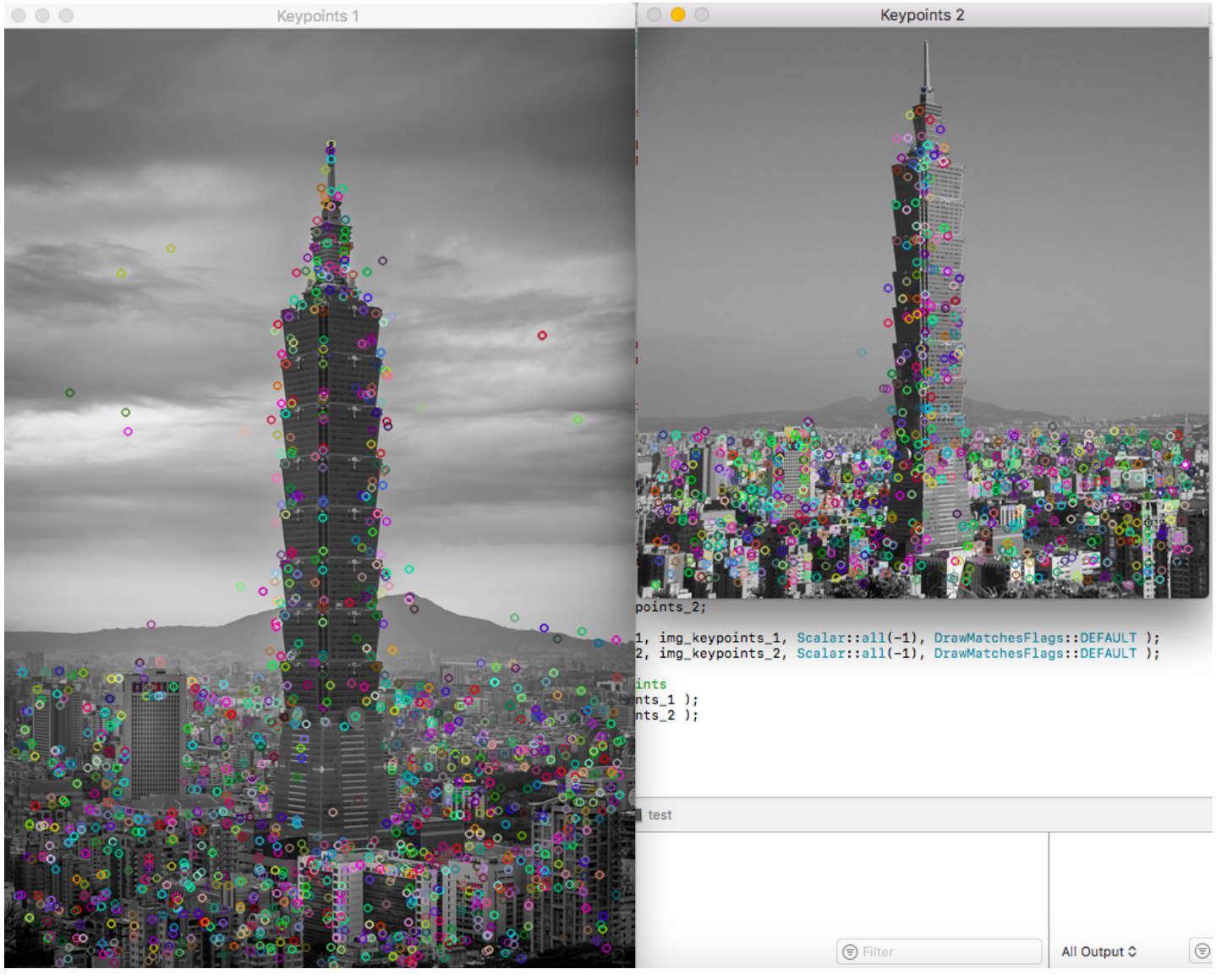
一个500*500的图像，得到约2000个特征

b）关键点描述子

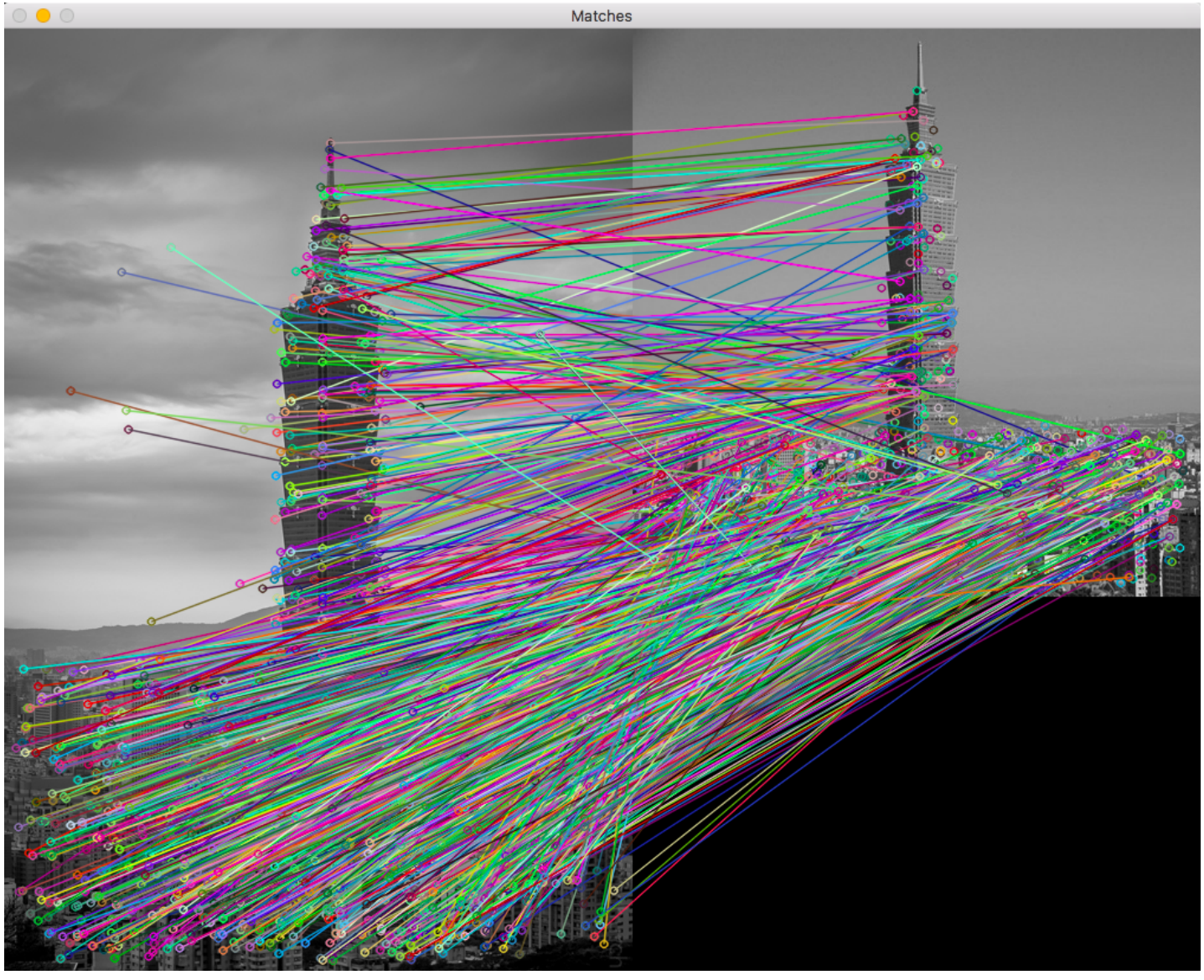
- 为关键点建立描述子向量
- 基于直方图，从而在不同光线与视角下能保持不变



SIFT特征提取结果


Keypoints 1
Keypoints 2
points_2;
1, img_keypoints_1, Scalar::all(-1), DrawMatchesFlags::DEFAULT);
2, img_keypoints_2, Scalar::all(-1), DrawMatchesFlags::DEFAULT);
ints
nts_1);
nts_2);
test
Filter
All Output

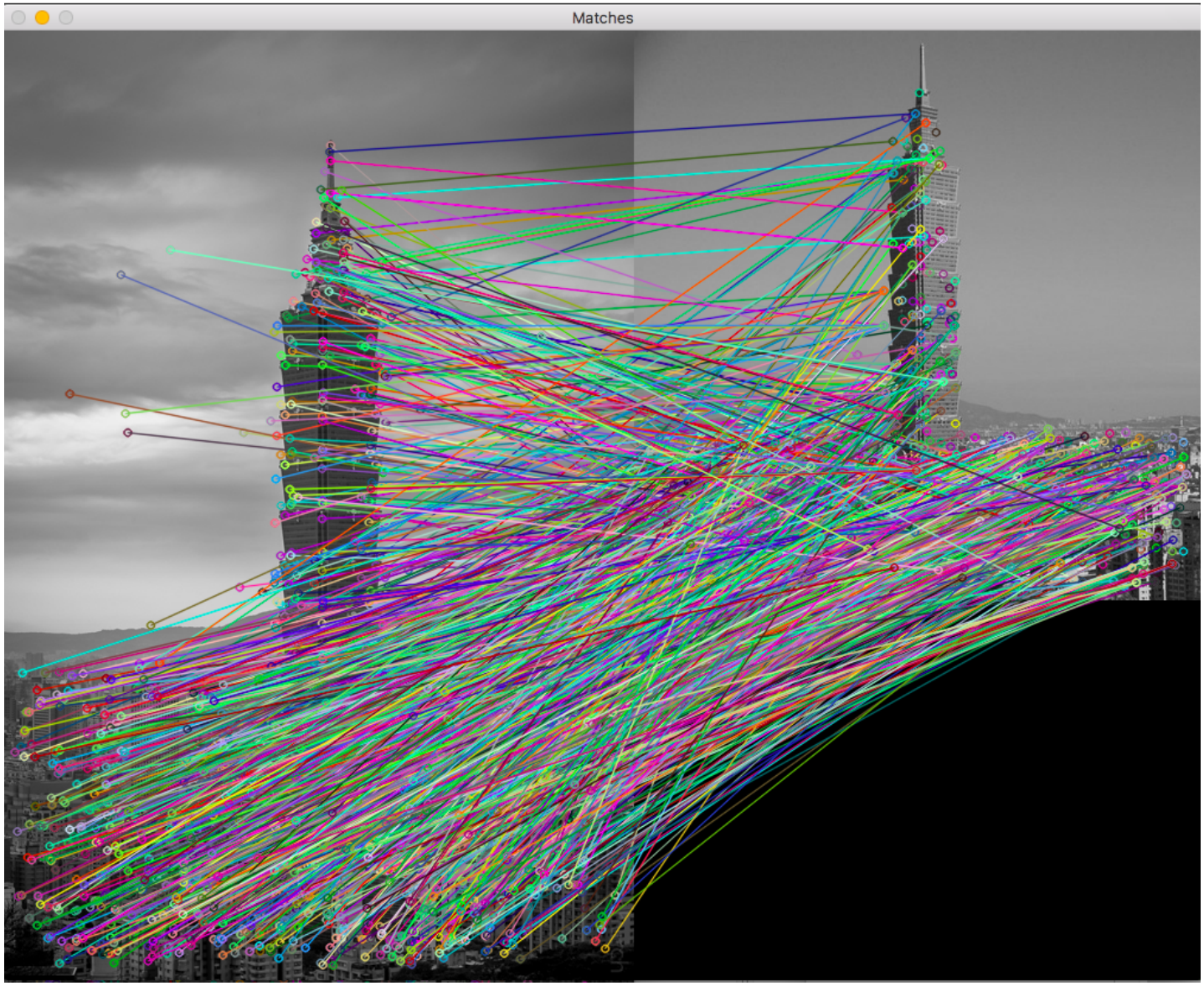
匹配



SURF

- Speeded Up Robust Features
- 加速稳健特征
- 2006年ECCV大会上发表
- 受SIFT启发，类似，速度更快，性能更稳定
  - 特征点侦测
  - 特征点邻近描述
  - 描述子配对

SURF算法效果



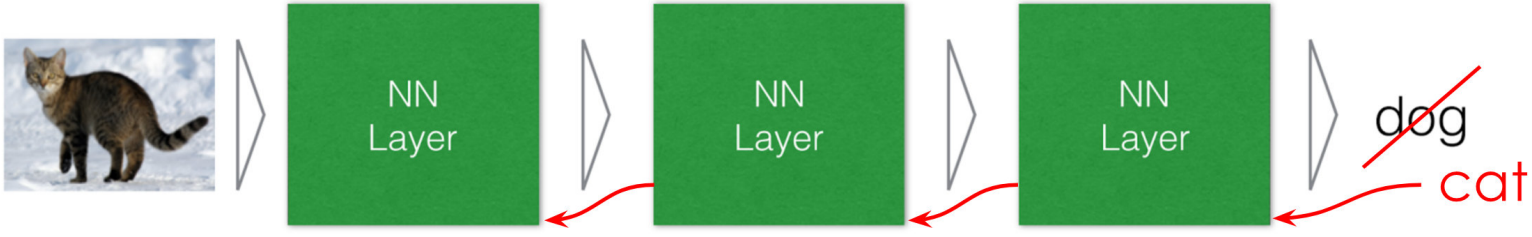
应用

- 物体辨识
- 机器人地图感知与导航
- 影像缝合
- 3D模型建立
- 手势辨识
- 影像追踪和动作比对

深度学习方法

深度卷积神经元网络

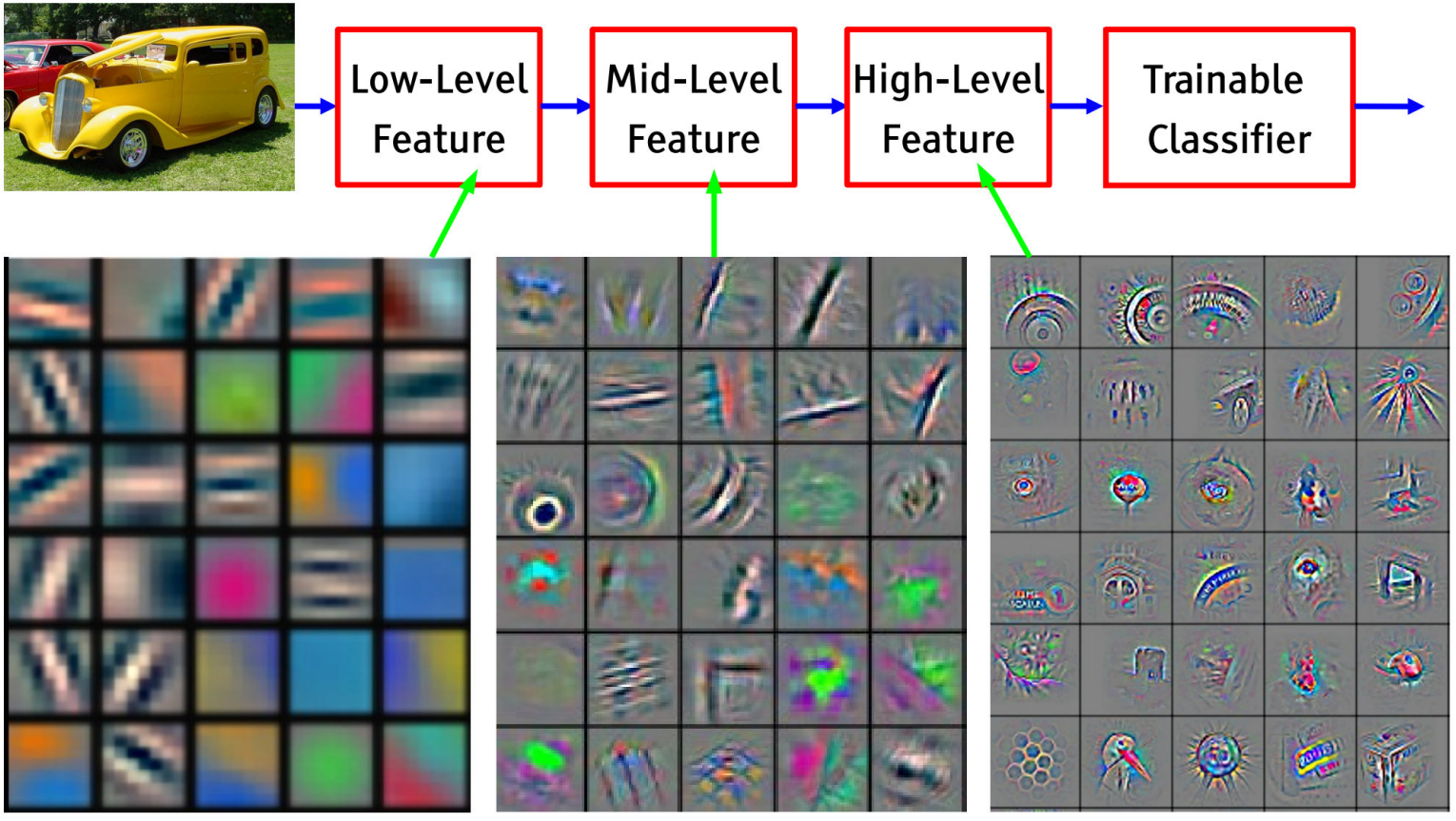
- 将原始数据直接送入多层神经元网络进行学习
- 多次卷积池化
- 出现错误，一路调整卷积核



Supervised
Learning

对各层卷积核的理解

- 底层提取简单特征，高层提取复杂特征



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

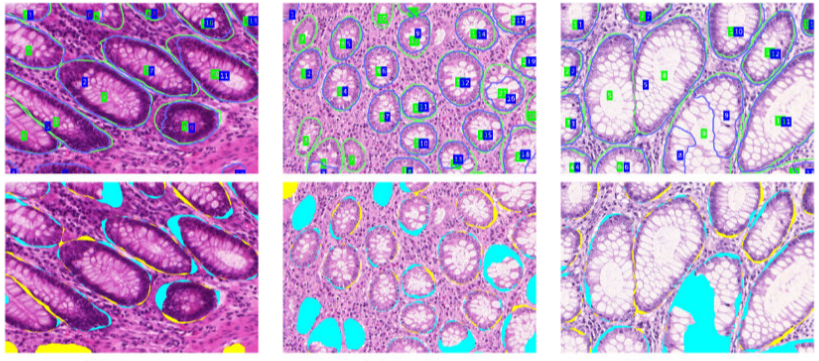
应用

目标检测，识别，手写体识别，目标分割

应用

疾病识别、人脸识别、脸部元素识别

[Stanford 2017]



[Nvidia Dev Blog 2017]

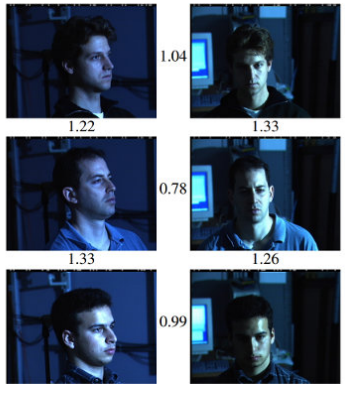
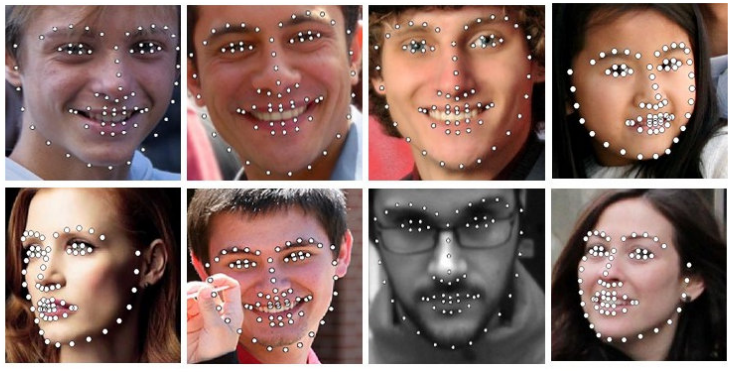


Figure 1. **Illumination and Pose invariance.**

[FaceNet - Google 2015]

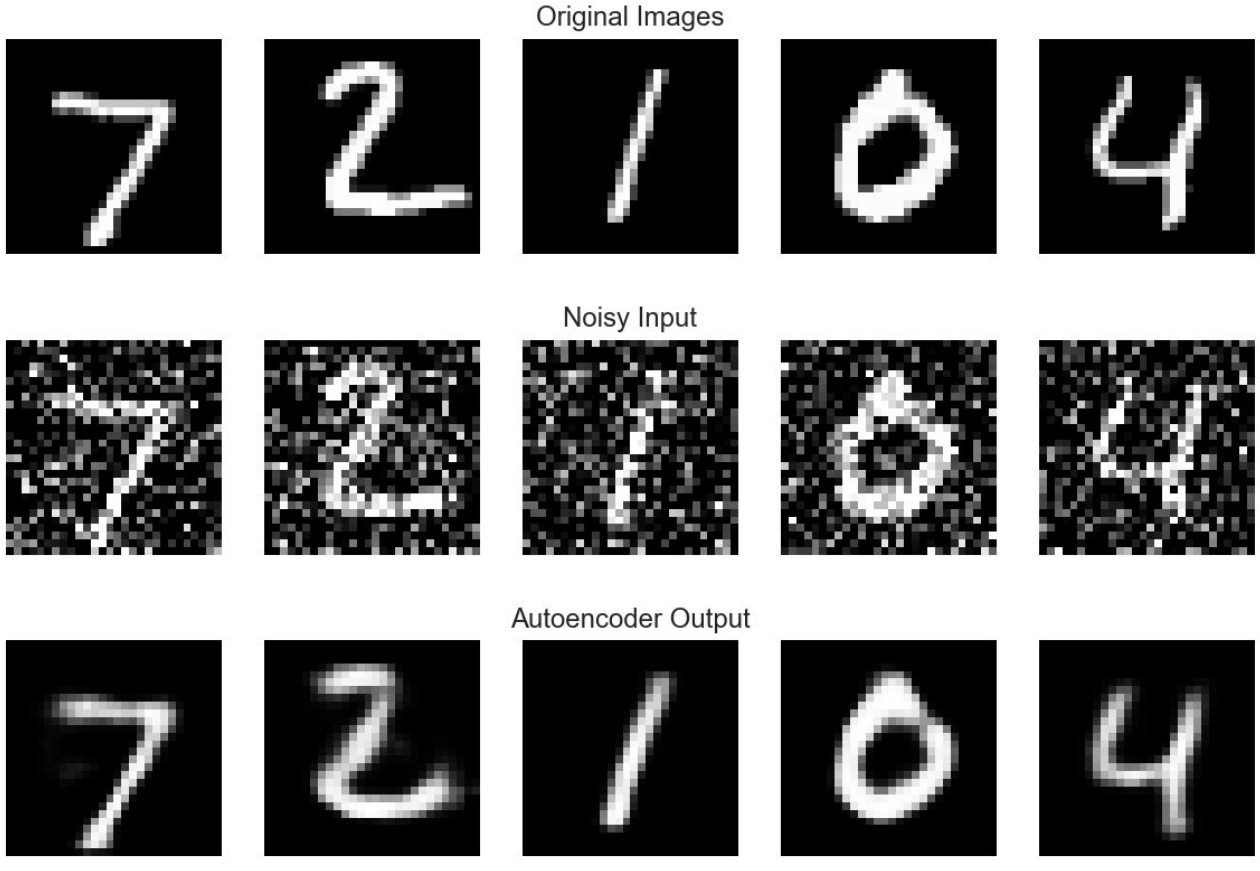


[Facial landmark detection CUHK 2014]

应用

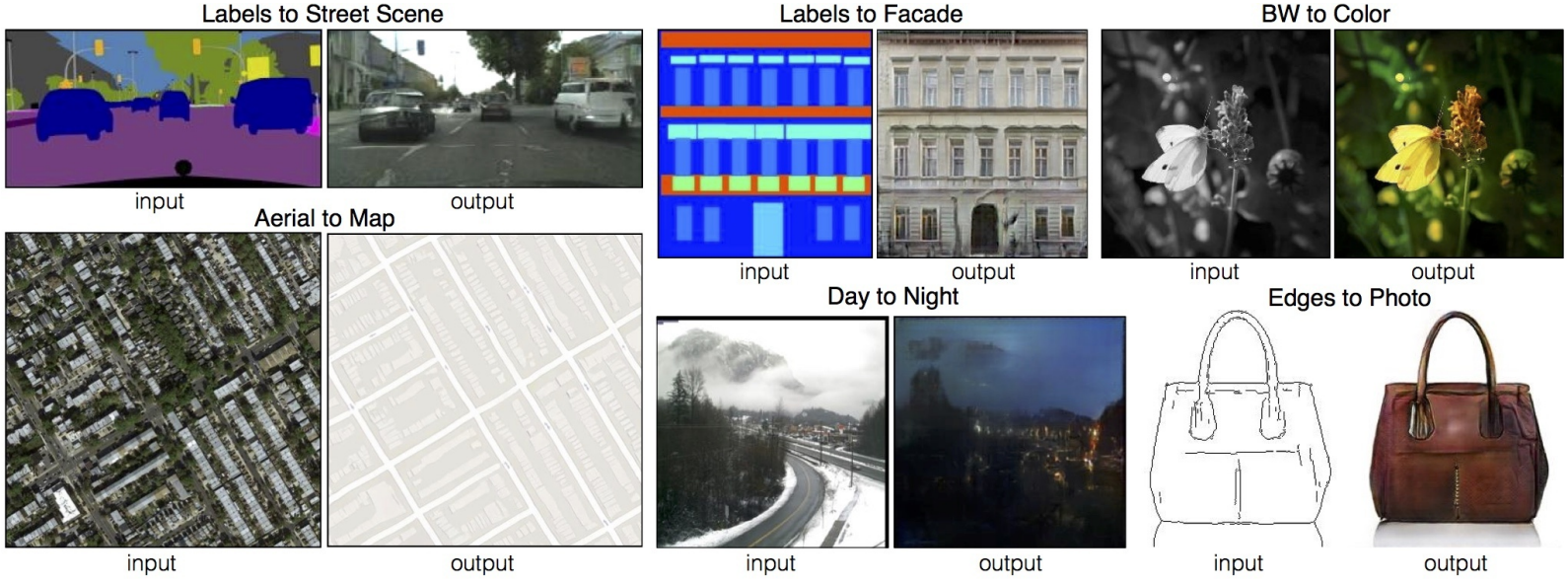
绘画、图像风格转换、清晰度增强

去噪



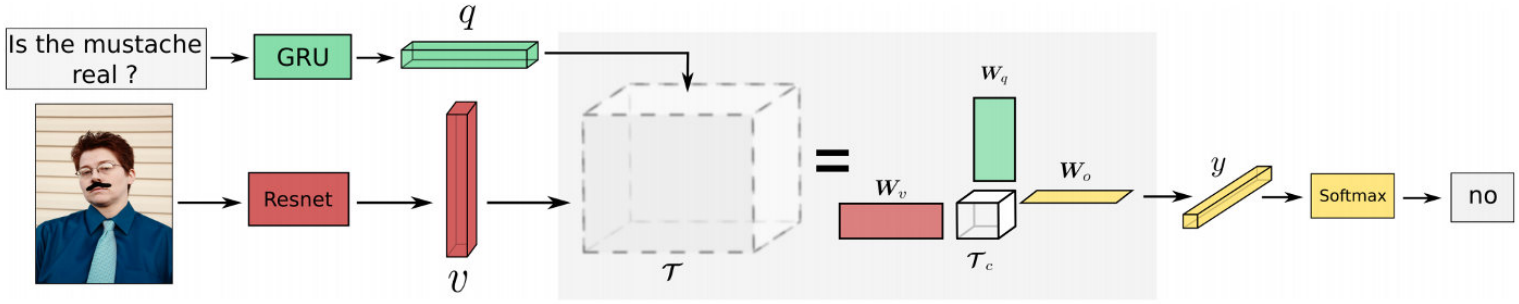
图像转换

- 图像还原、渲染、着色
- 地图提取、场景转换



图像理解

图像 - 问答 - 文本描述



[VQA - Mutan 2017]

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

[Karpathy 2015]

实时图像理解（2015）

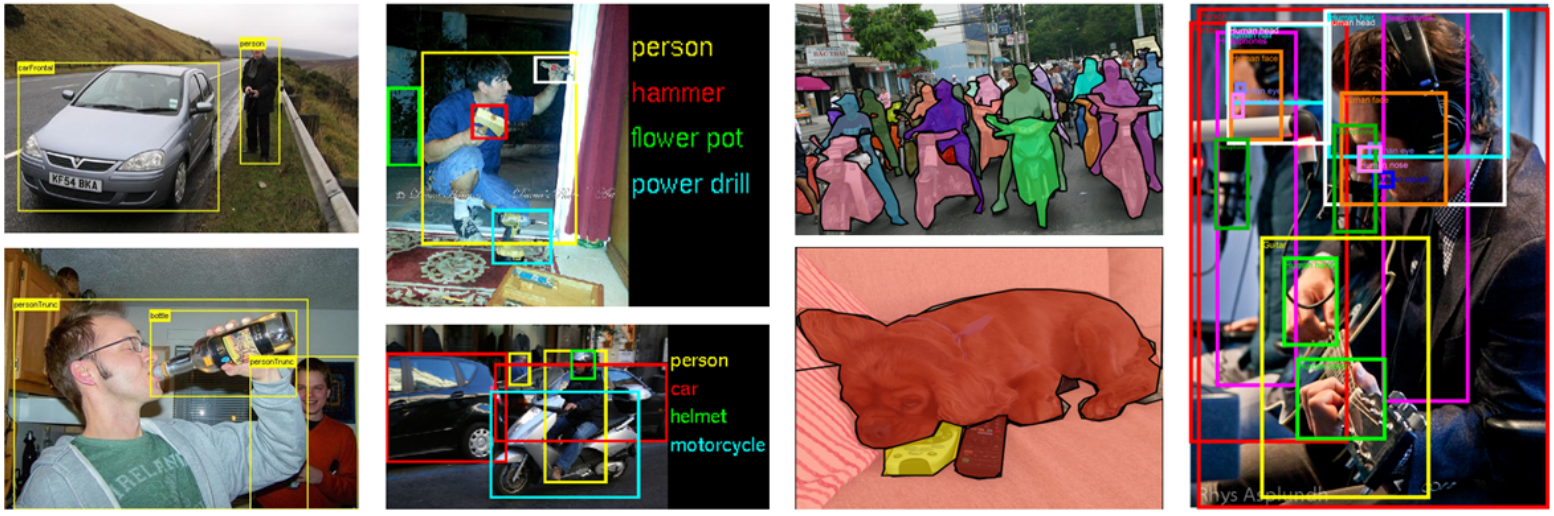
人工智能：实时场景理解，文本生成

目标检测和识别

Object Detection & Recognition

基本研究问题

目标检测、分割、识别



(a) (b) (c) (d)

困难

遮蔽、干扰、噪声

拖把狗

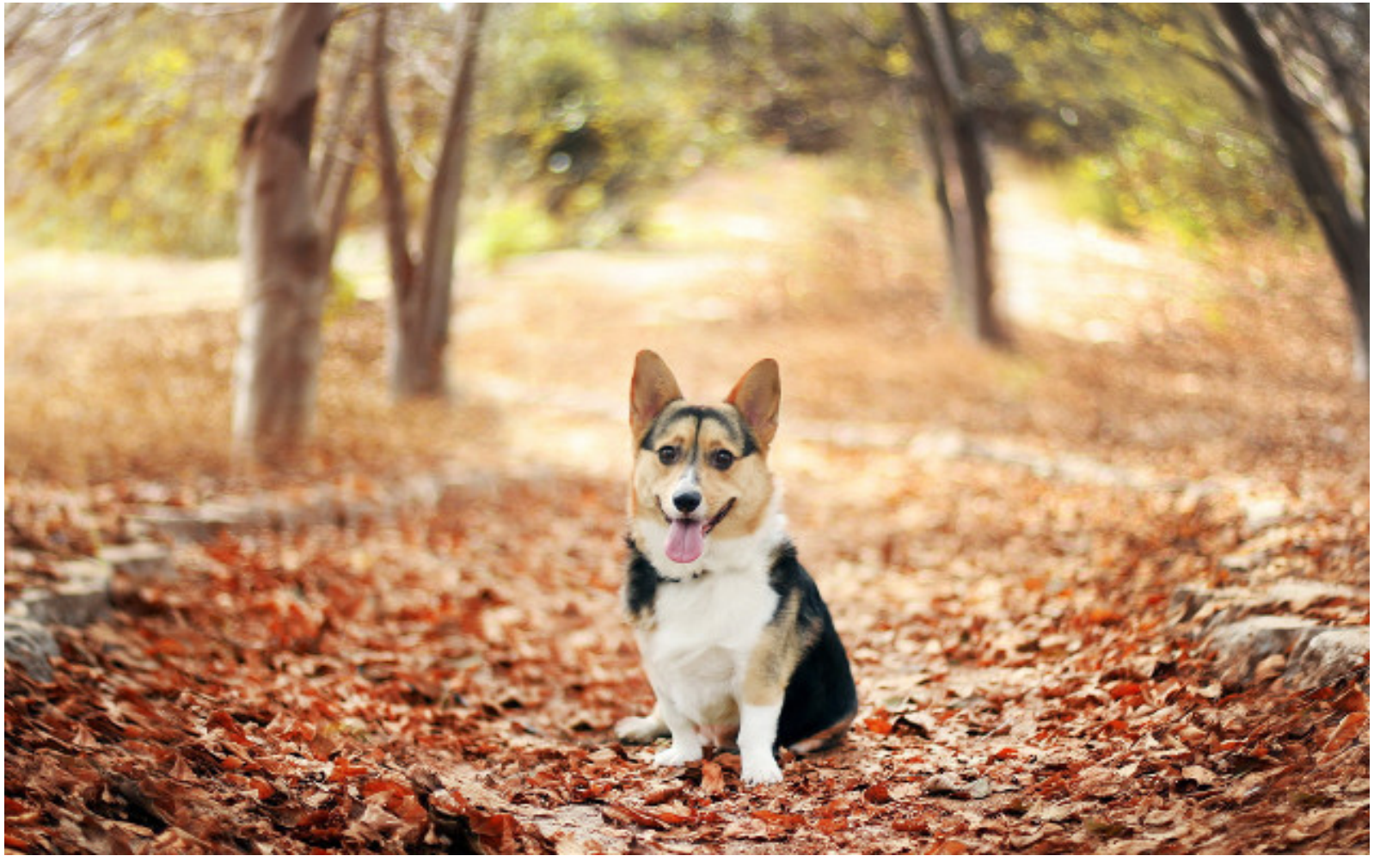
蛋糕狗

云雾

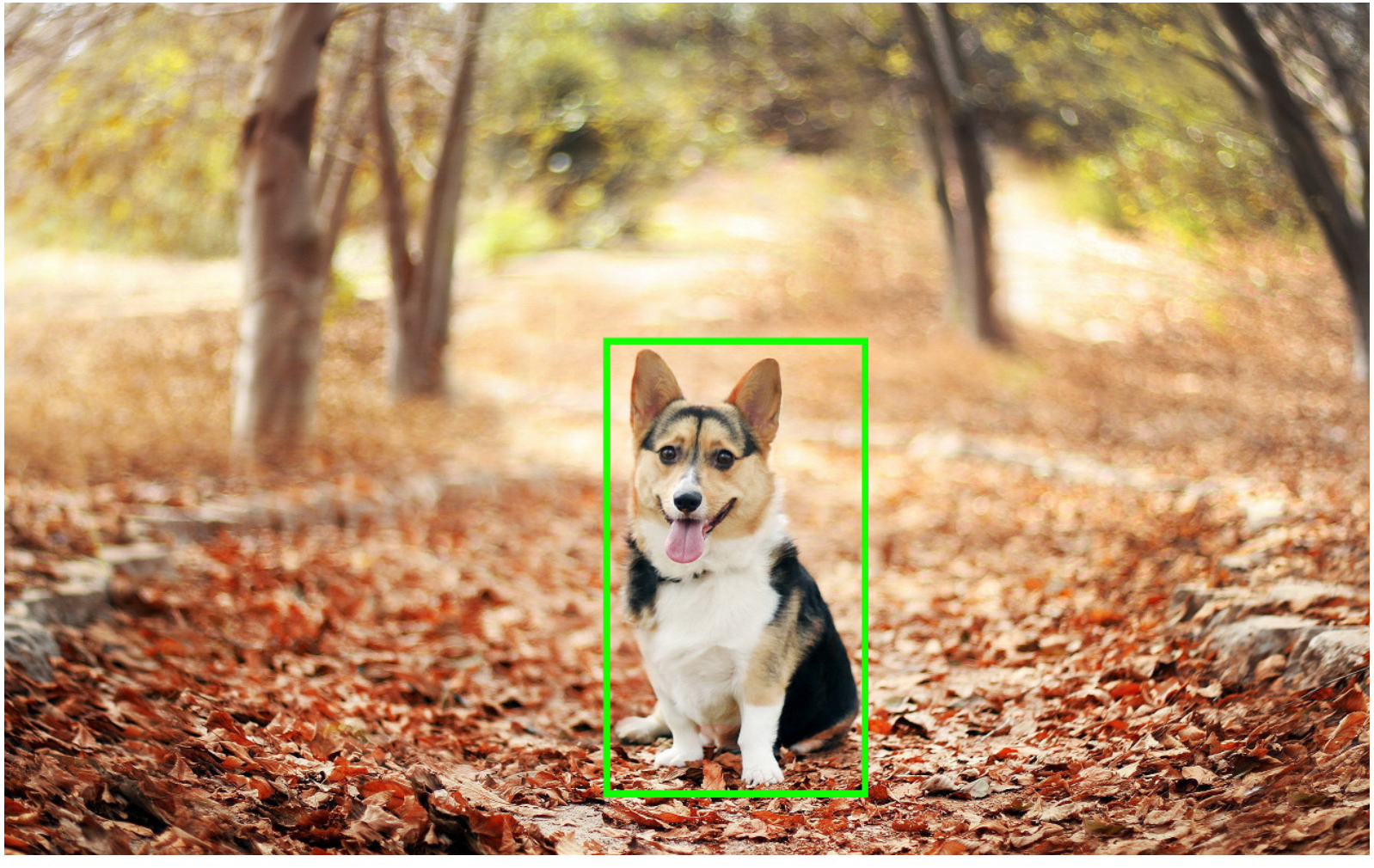
目标检测

Object Detection

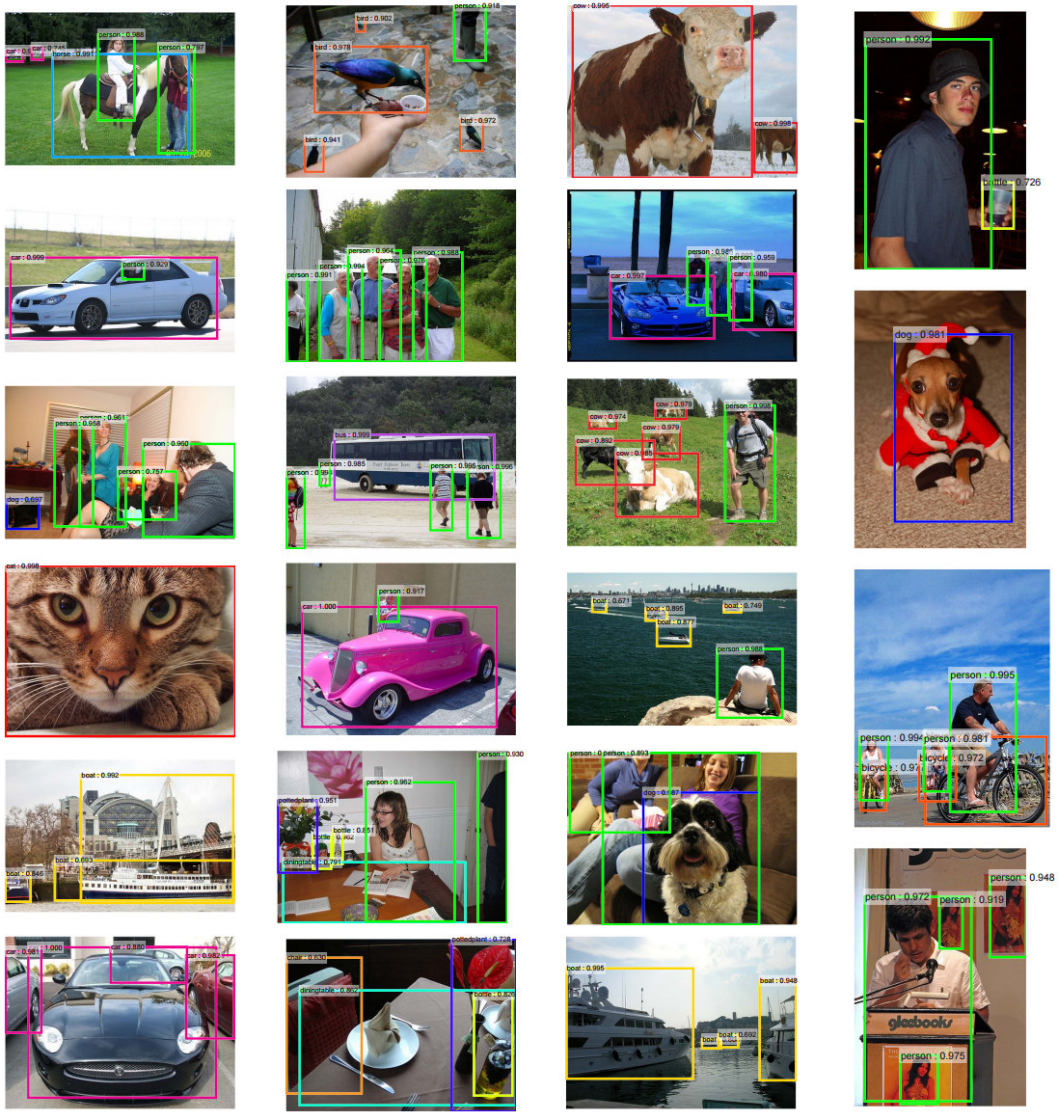
图像



目标检测



目标检测

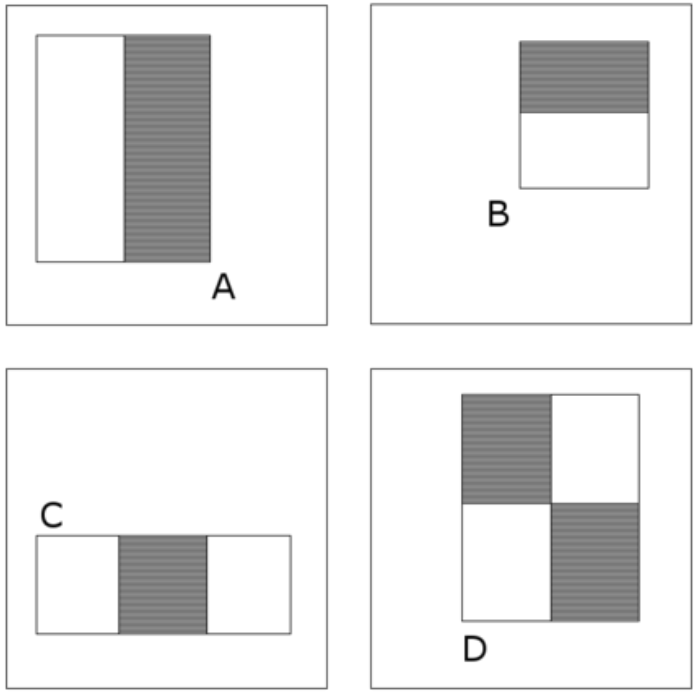


1）传统方法

- V-J检测
- HOG检测
- DPM算法

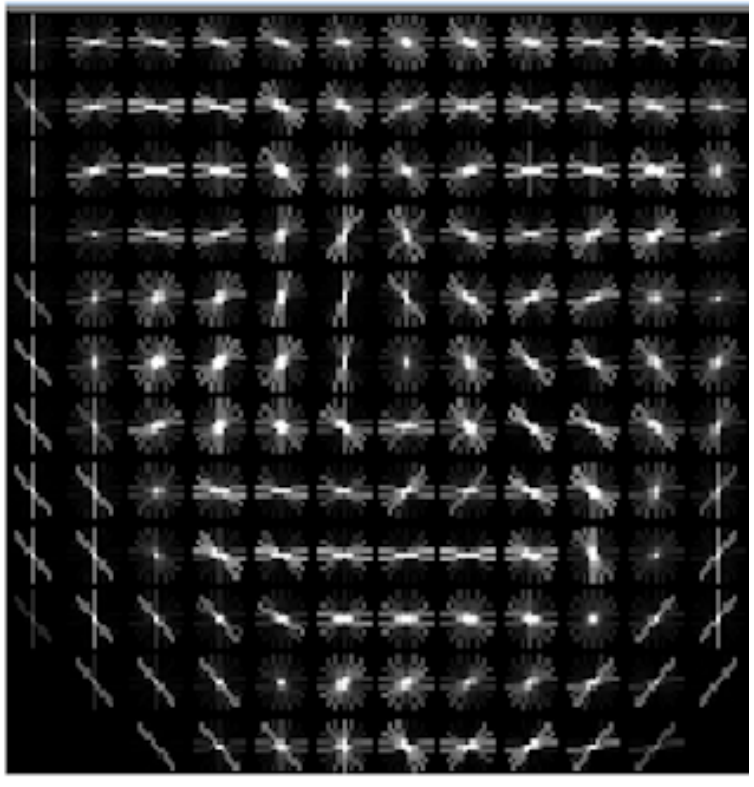
V-J检测

- 2001年，Paul Viola，Michael Jones提出
- 人的正脸检测
- Haar特征



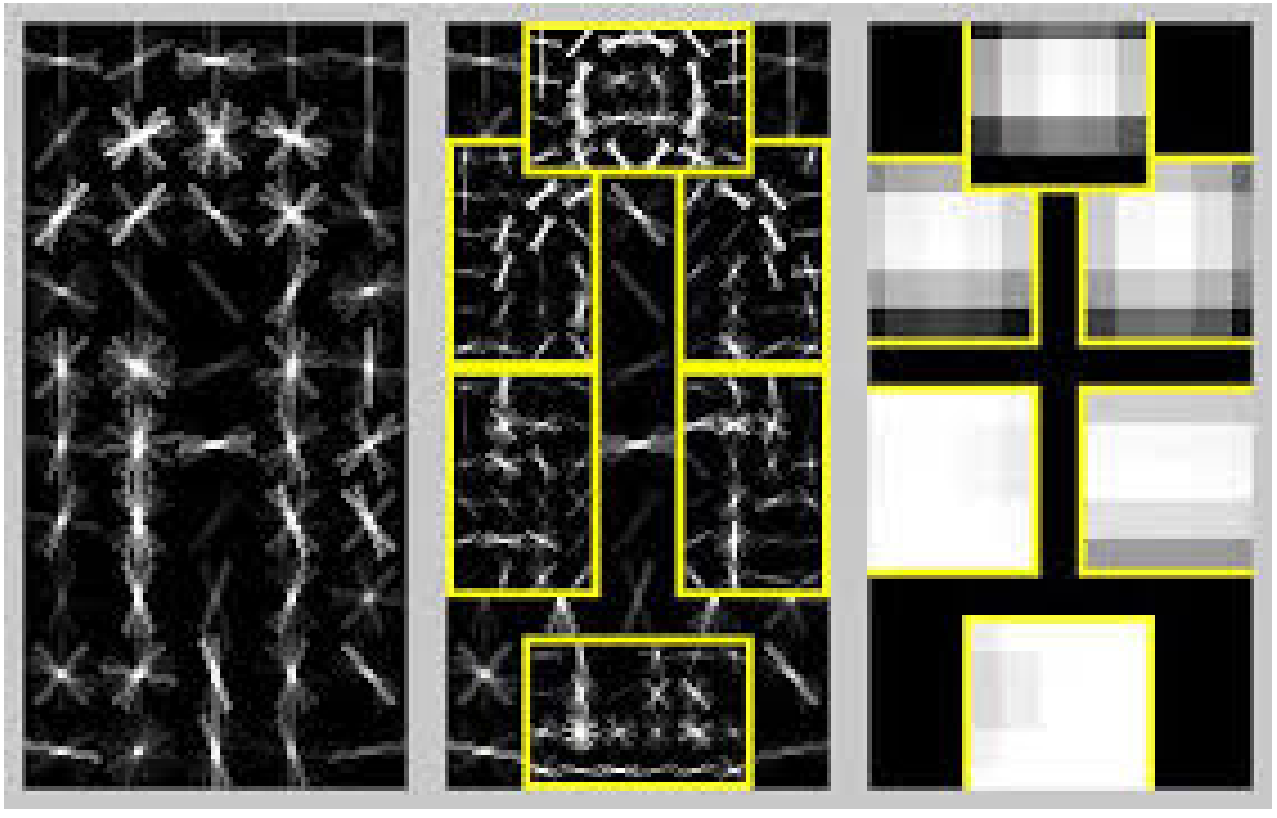
HOG检测

像素梯度



DPM算法

- Deformable Part-Based Model
- 各部分有自己的分类器（如：眼睛、嘴）
- 各部分位置应该合理（如：眼睛在嘴上面）

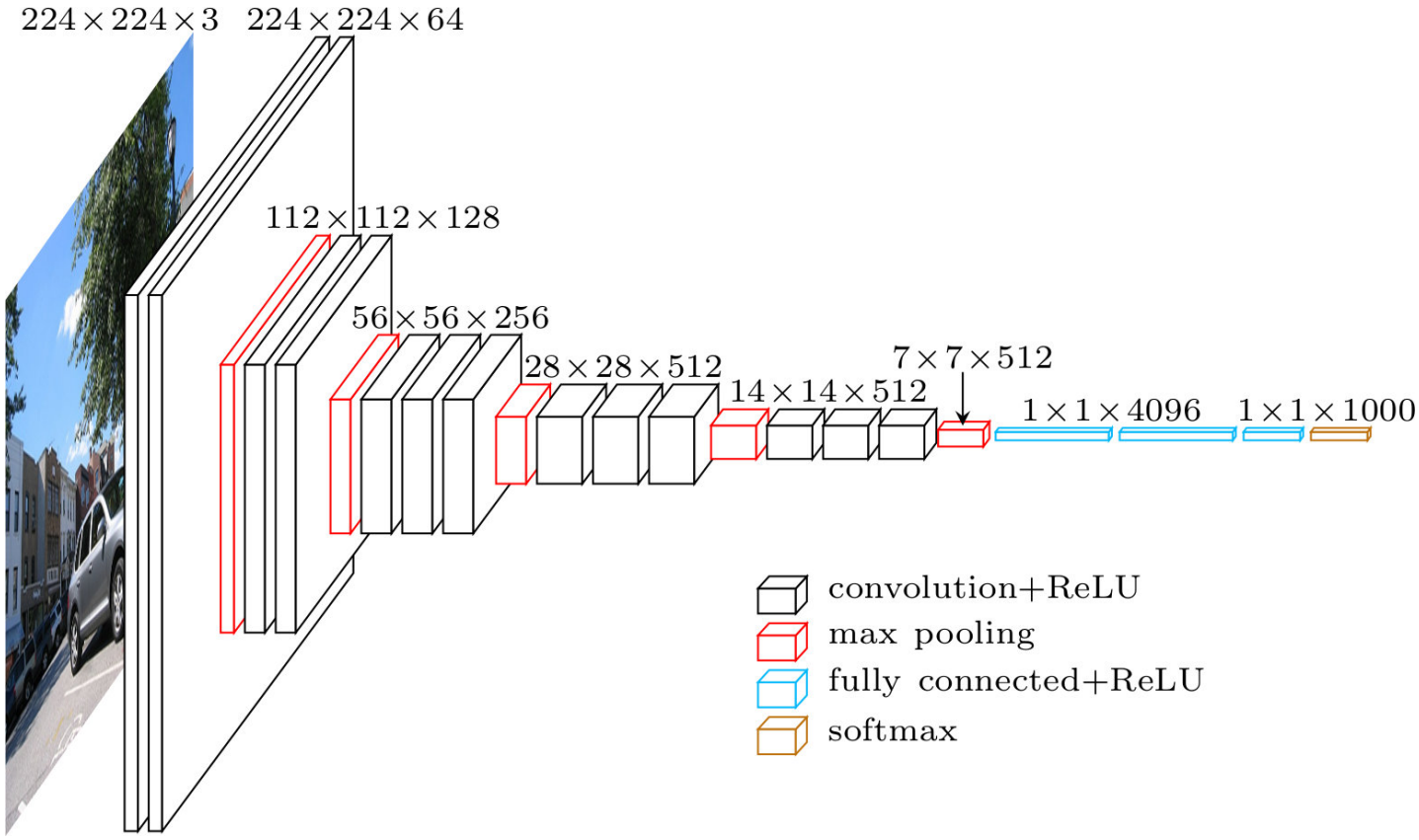


2）深度学习方法

- 2012年，AlexNet
- 双阶段检测器
  - 先找区域，再识别目标
  - RCNN、Pyramid Networks
- 单阶段检测器
  - 不找区域，直接识别目标
  - YOLO、SSD、Retina-Net
- 评估mAP
  - VOC 83% （2018），COCO（69% 2019）

VGG16

- CNN目标识别
- 牛津大学，K. Simonyan，A. Zisserman，2014年

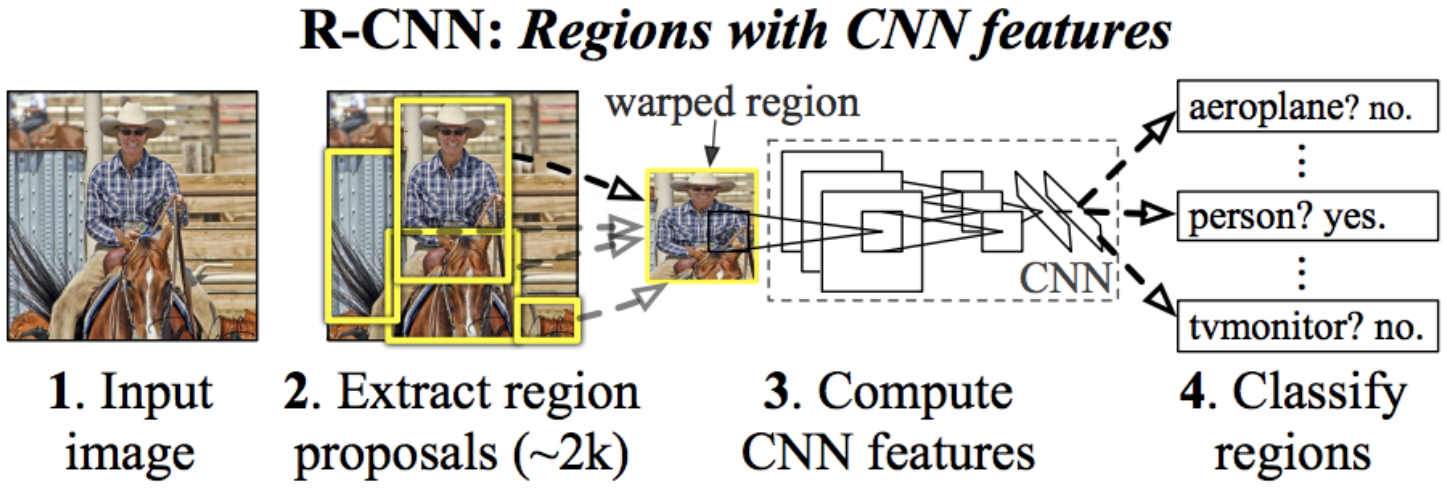


双阶段检测器

先找区域，再识别目标

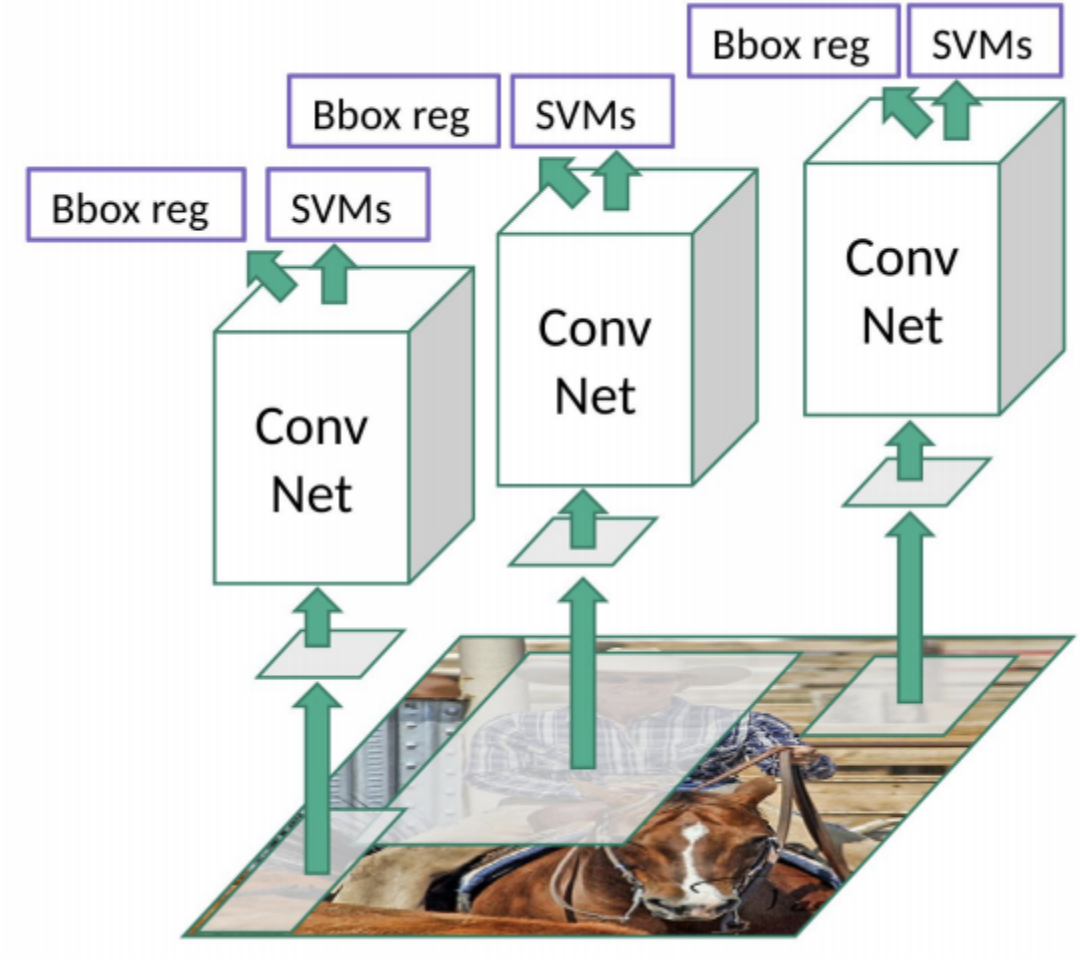
RCNN

- 初始化小区域
- 贪婪算法合并区域
- 最后选出2000个可能区域



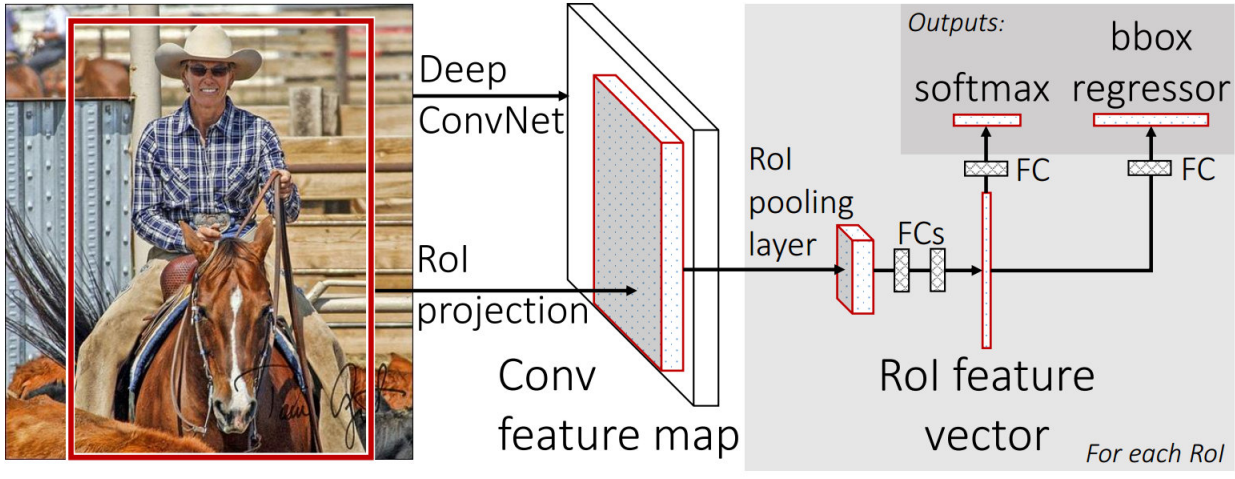
RCNN

- CNN除了目标识别，也建议调整区域



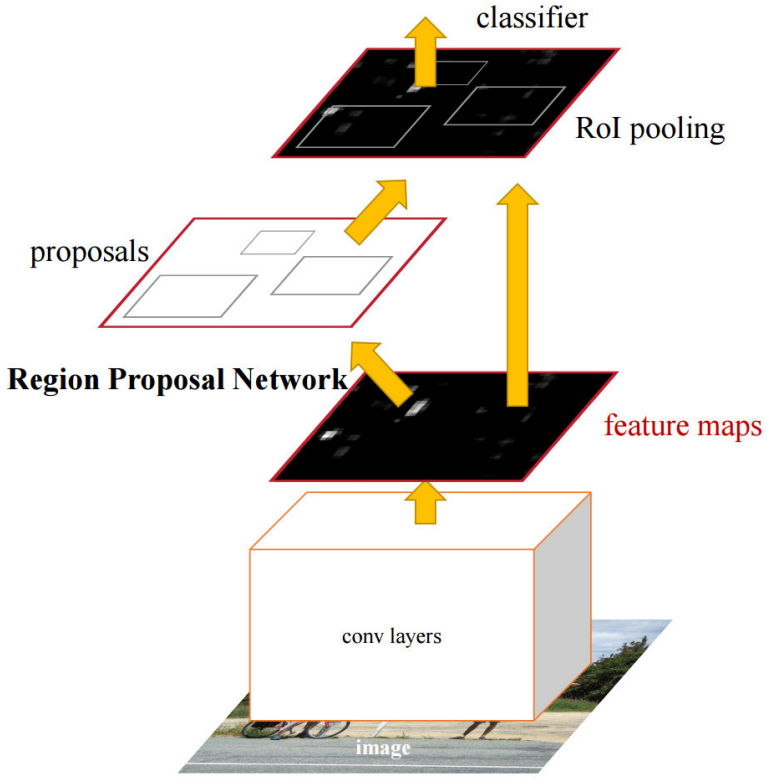
Fast R-CNN

- R-CNN要对2000个区域分别CNN
- 改进
  - 对全部图像做一次CNN
  - 在得到的特征地图上，选出可能区域
- 速度提高几十倍



Faster R-CNN

- 去掉选择性搜索可能区域这一费时工作
- 用另外一个网络预测可能出现目标的区域
- 速度又提高10倍

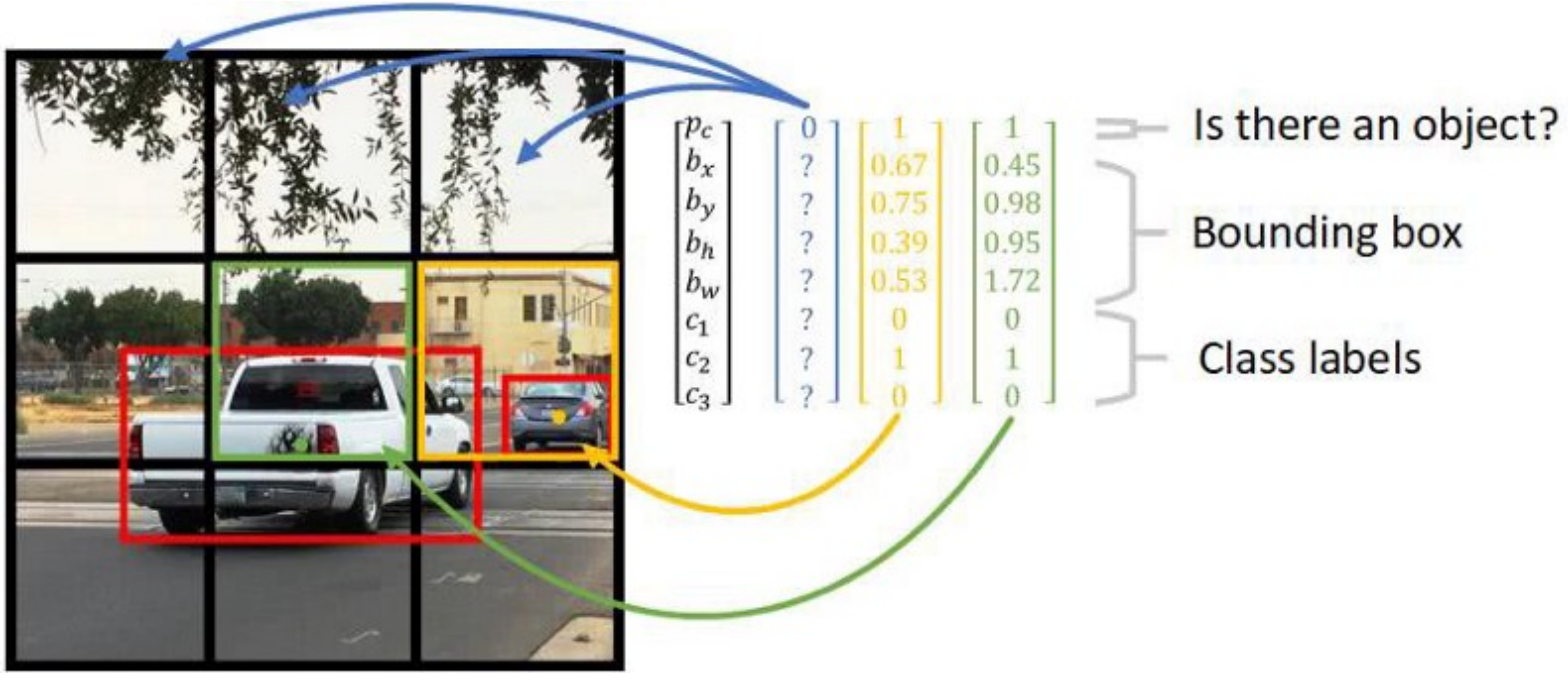


单阶段检测器

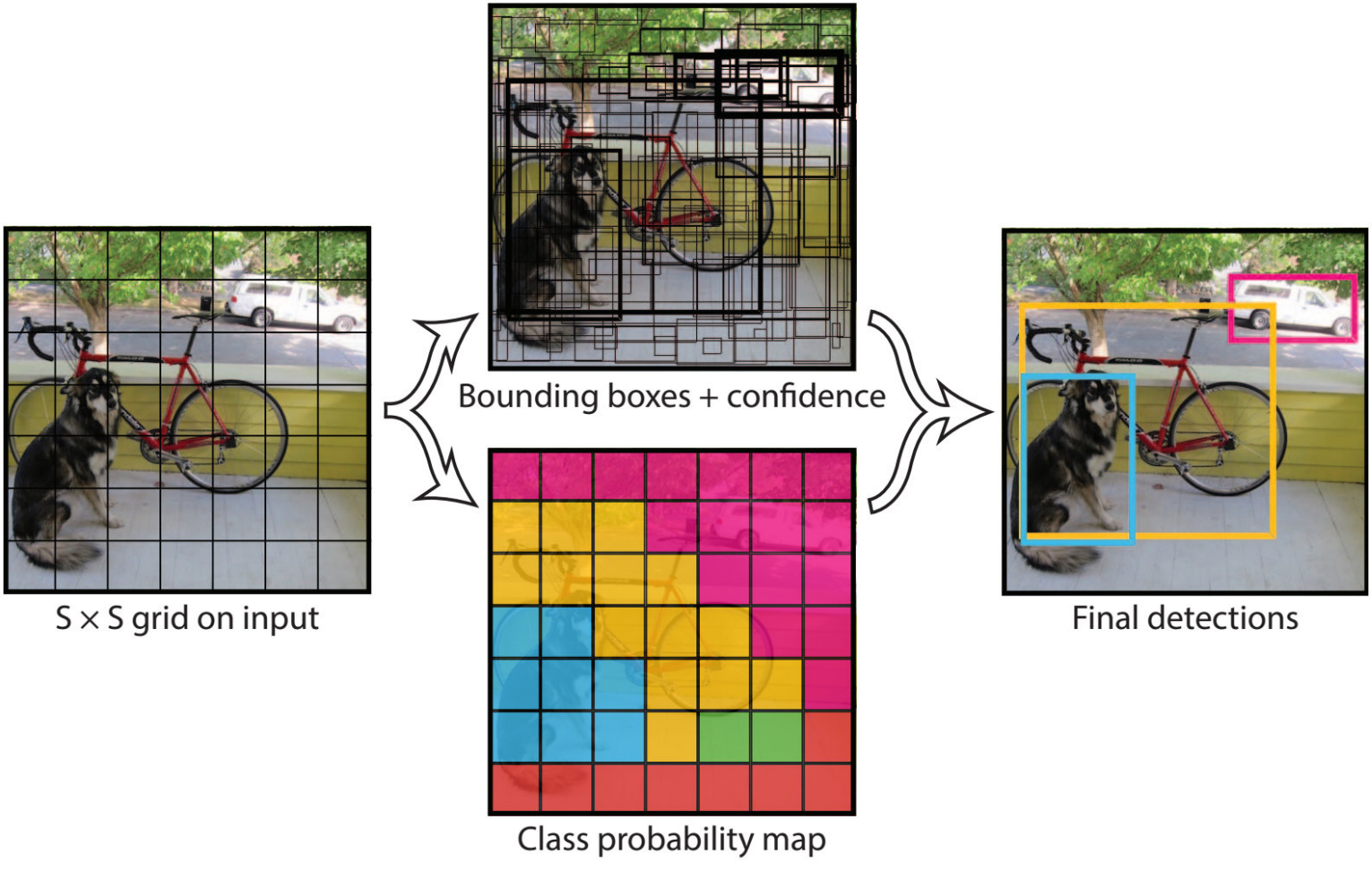
不找区域，直接识别目标

YOLO

- You Only Look Once（只看一次），2015年提出
- 图像分成小块，每块选多个可能的目标区域。
- 对每个区域，卷积网络给出目标区域的偏移建议和目标类型判断



效果


Bounding boxes + confidence
S × S grid on input
Class probability map
Final detections

YOLO

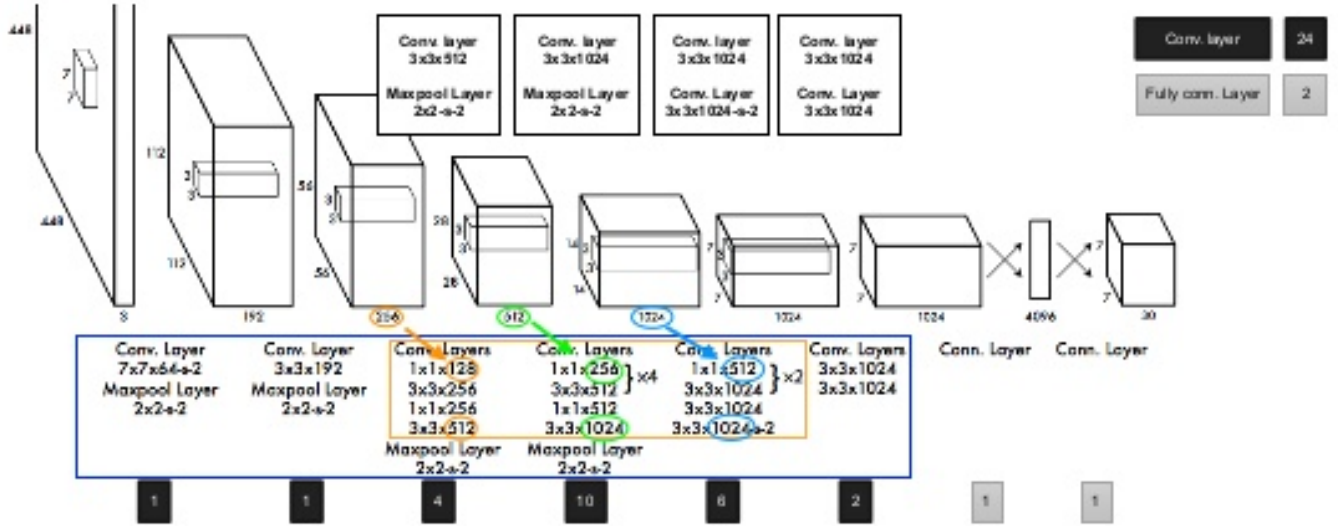
- 网络参考GoogleNet
- 速度更快，每秒45帧没问题

Appendix: GoogLeNet

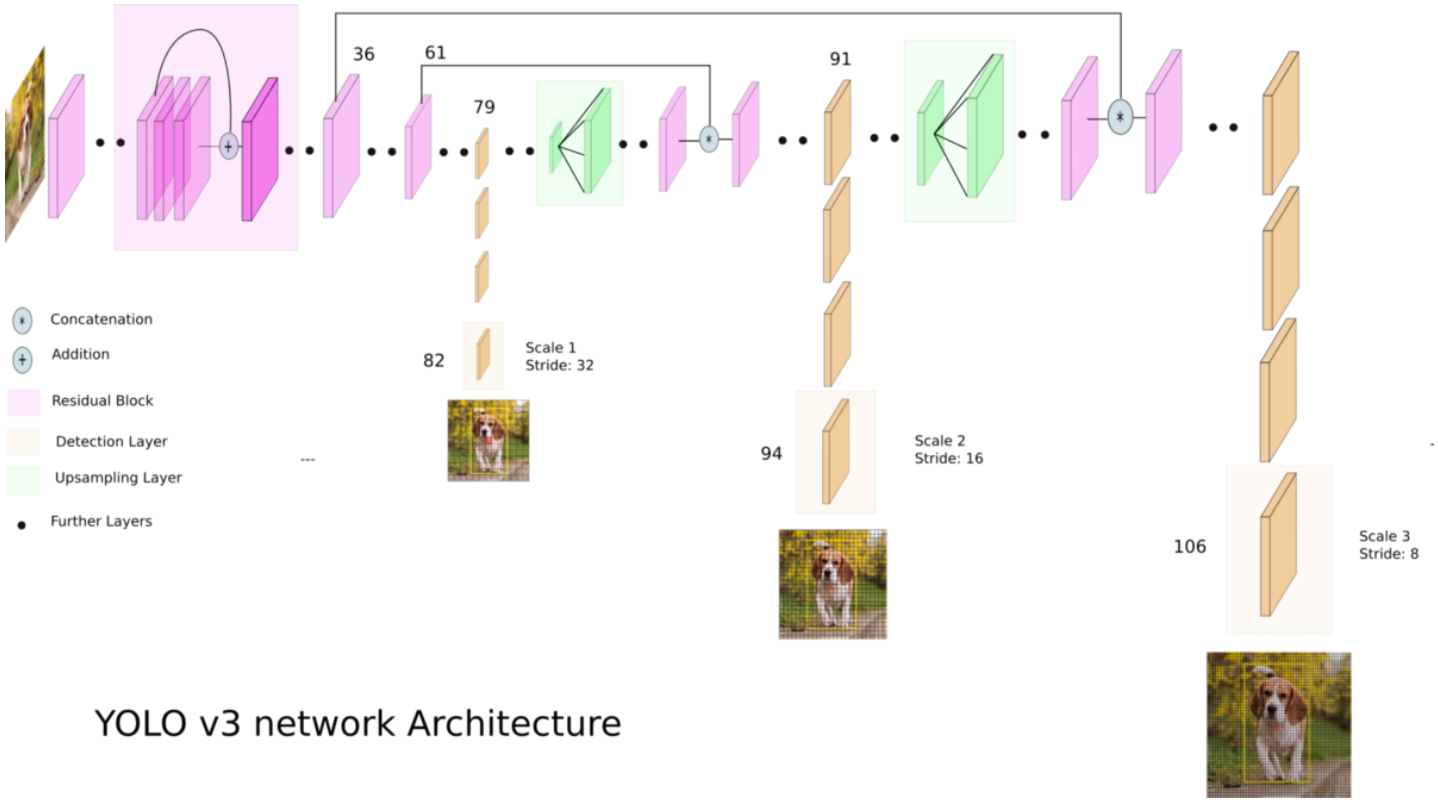
Network Design: YOLO

- Modified GoogLeNet
- 1x1 reduction layer ("Network in Network")

Our network architecture is inspired by the GoogLeNet model for image classification [34]. Our network has 24 convolutional layers followed by 2 fully connected layers. Instead of the inception modules used by GoogLeNet, we simply use 1×1 reduction layers followed by 3×3 convolutional layers, similar to Lin et al [22]. The full network is shown in Figure 3.



YOLO v3

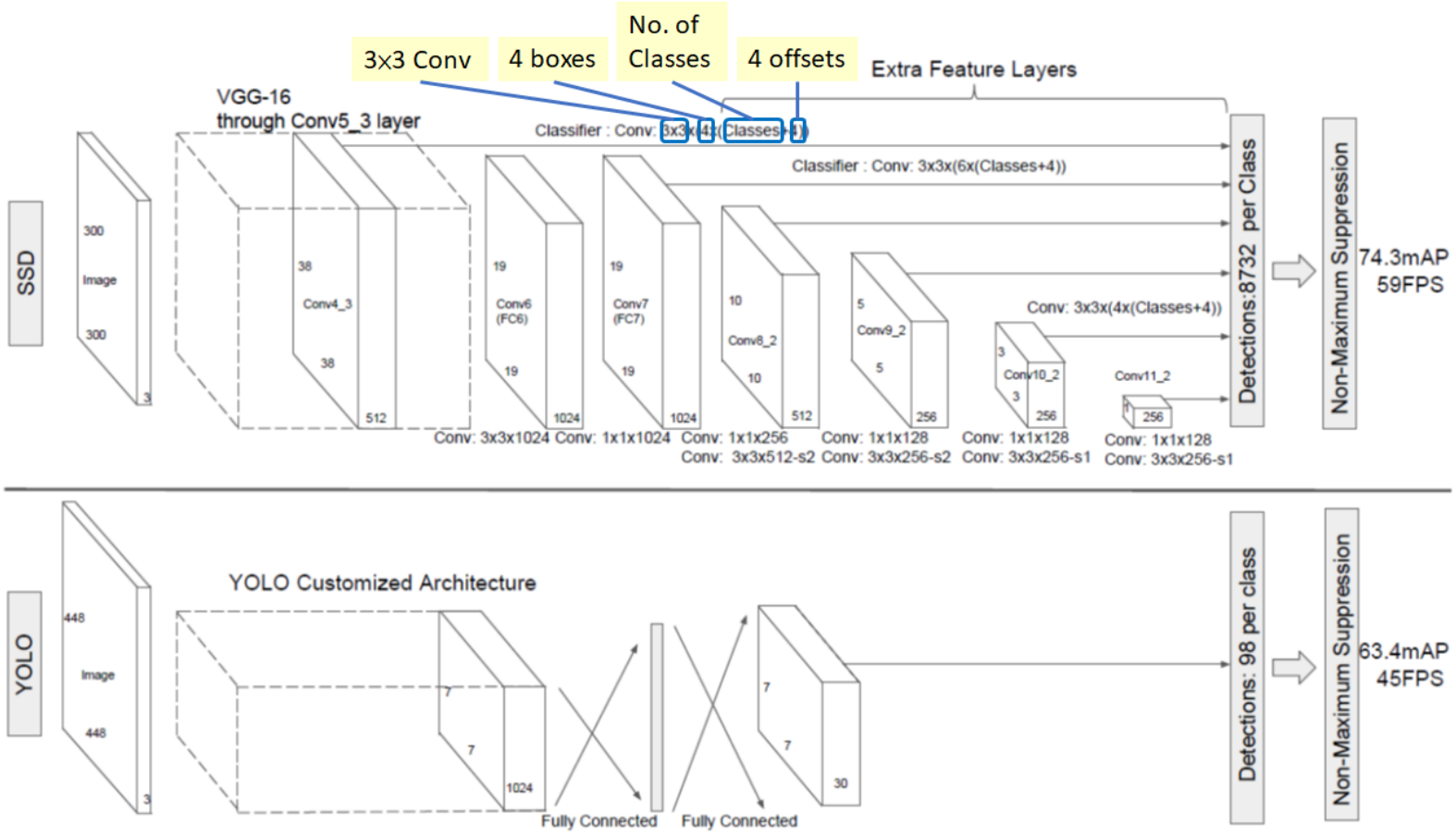


YOLO v3 network Architecture

- 更准确
- 缺点：小目标识别困难，比如鸟群

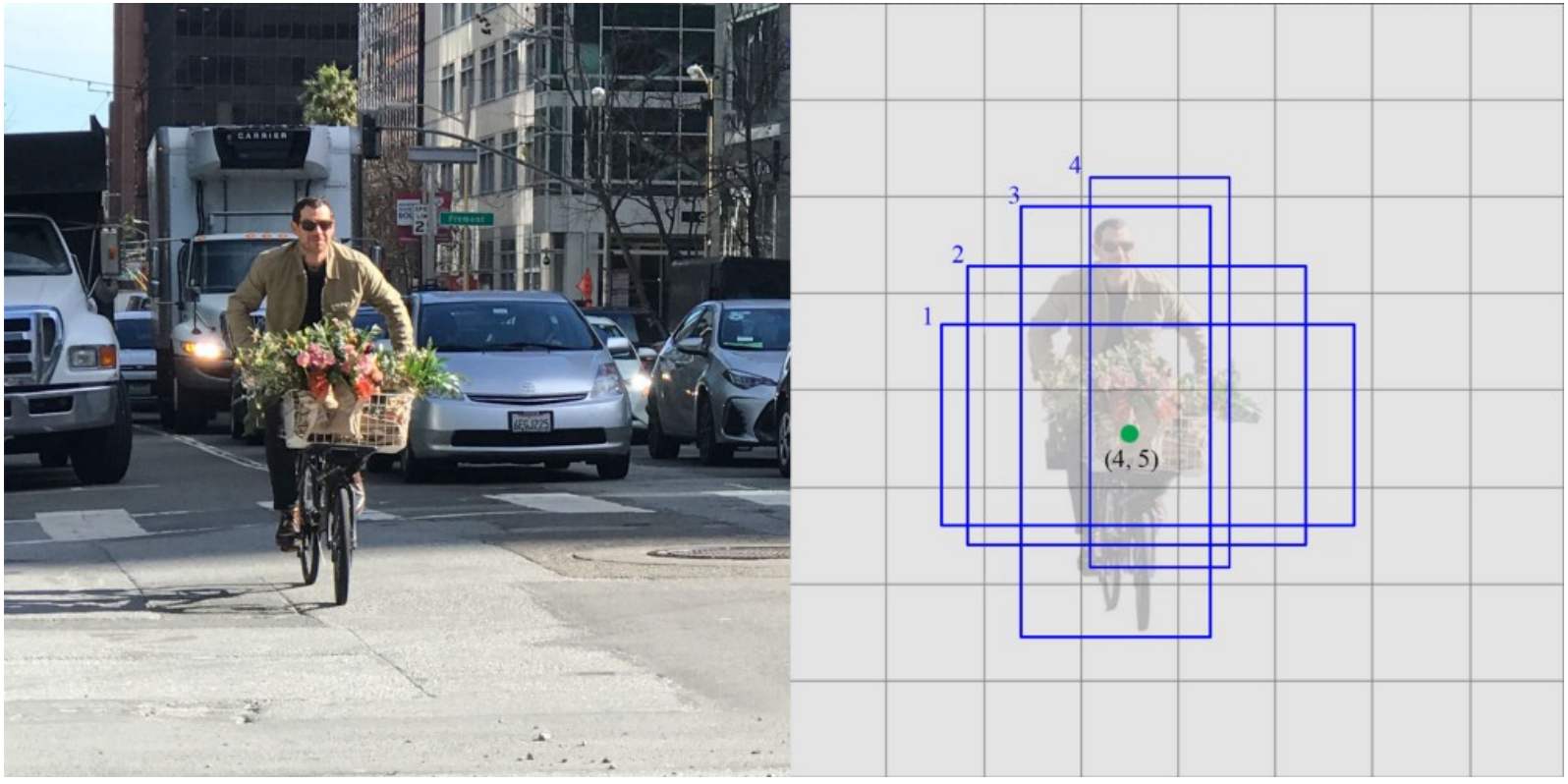
SSD

- Single Shot MultiBox Detector
- 2016年ECCV会议上提出



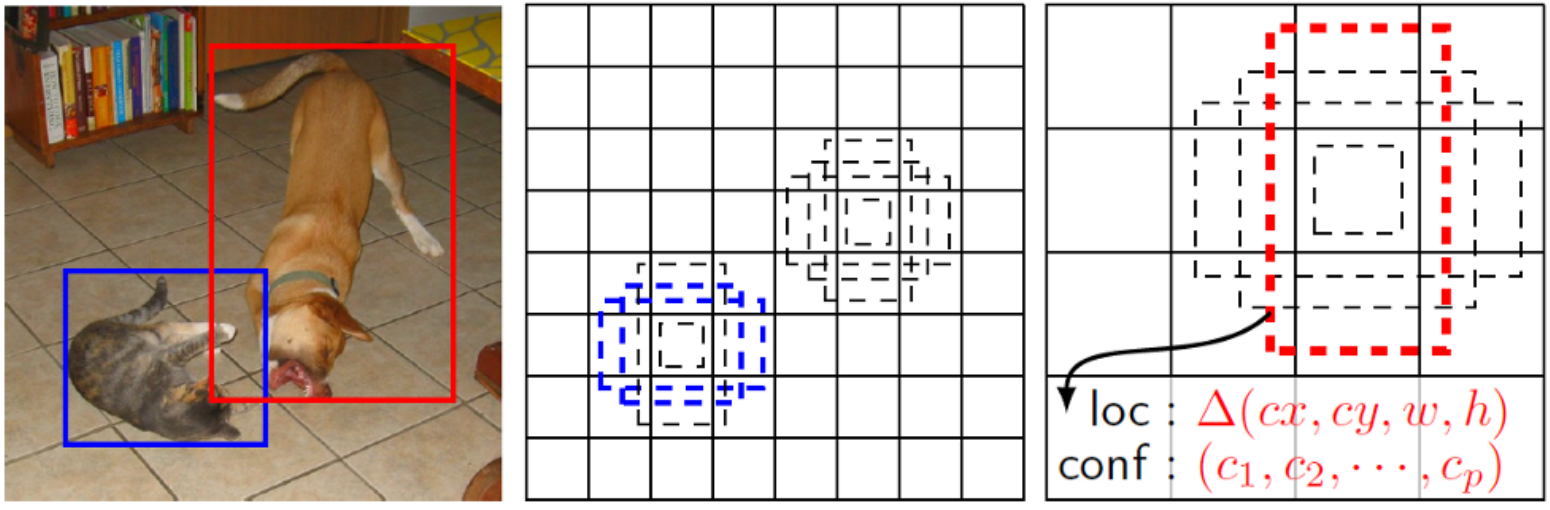
默认目标盒形状

- 汽车，人有特定形状
- 手工选择初始四种默认目标盒子



多尺度特征地图

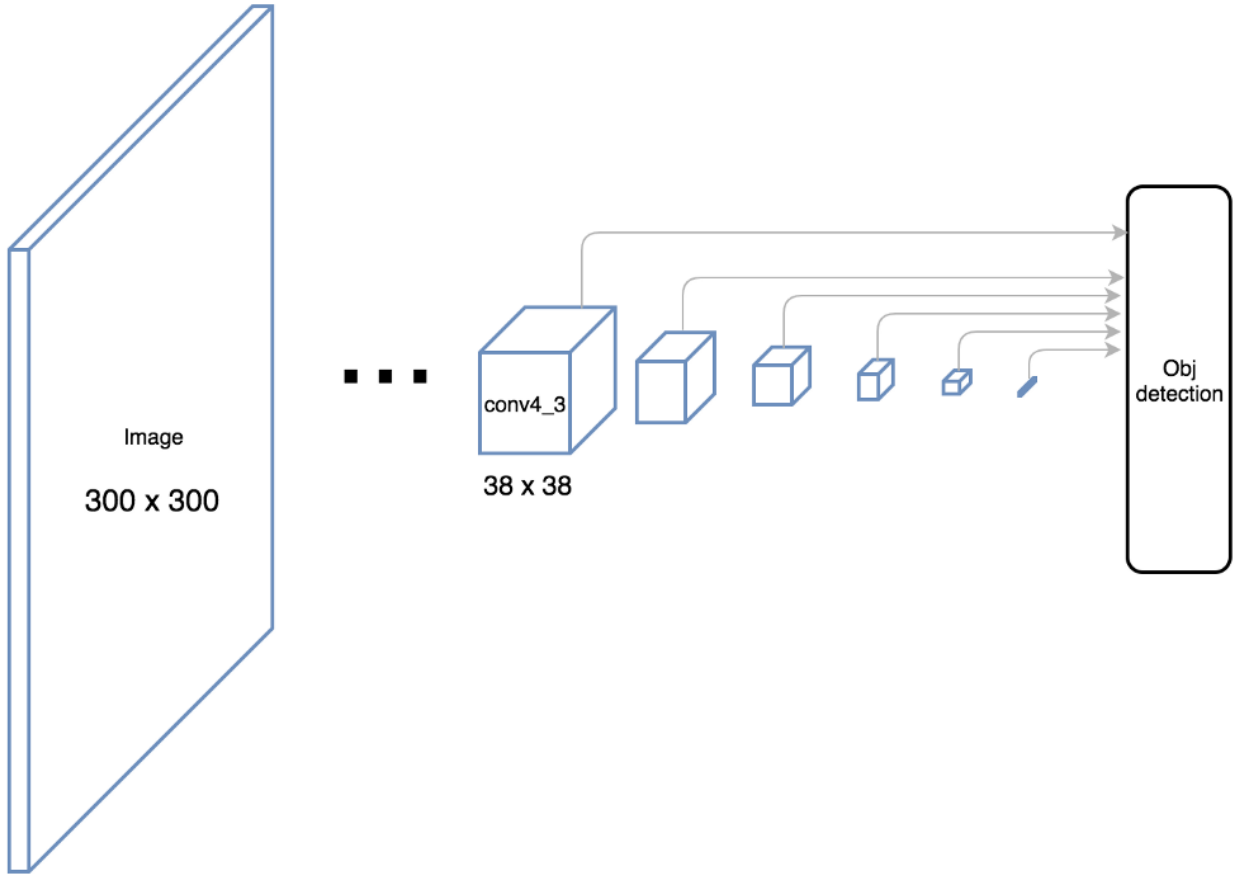
- 分不同尺度的小块，检测不同尺度对象



(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

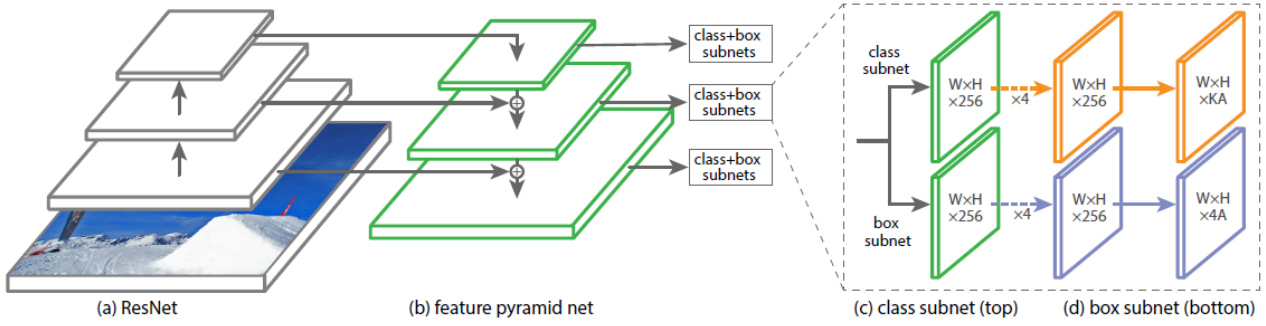
多分辨率卷积网络

- VGG后另加6个CNN，分辨率各不相同
- 高分辨CNN帮助识别小目标



RetinaNet

- 2017 ICCV
- 骨干网：ResNet + Feature Pyramid Net（FPN）特征金字塔网
  - 金字塔网每一级的分辨率不同
- 任务网
  - 目标识别
  - 边界盒发现

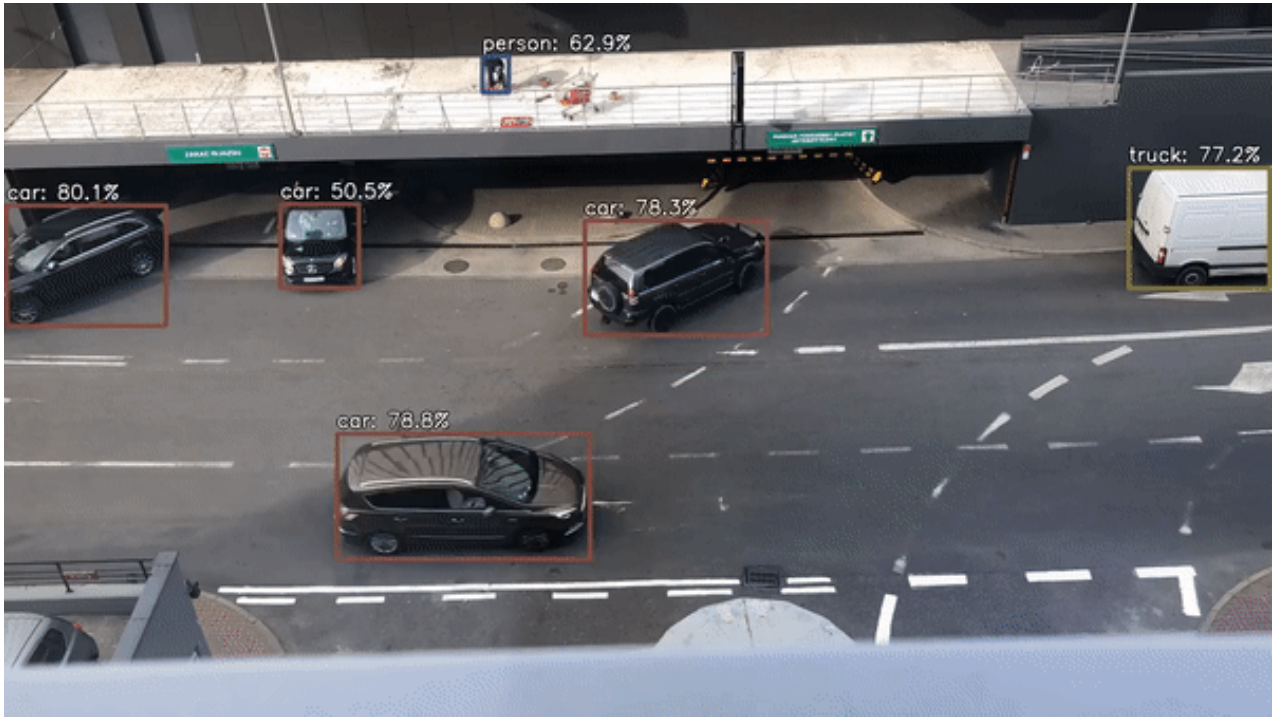


Focal Loss

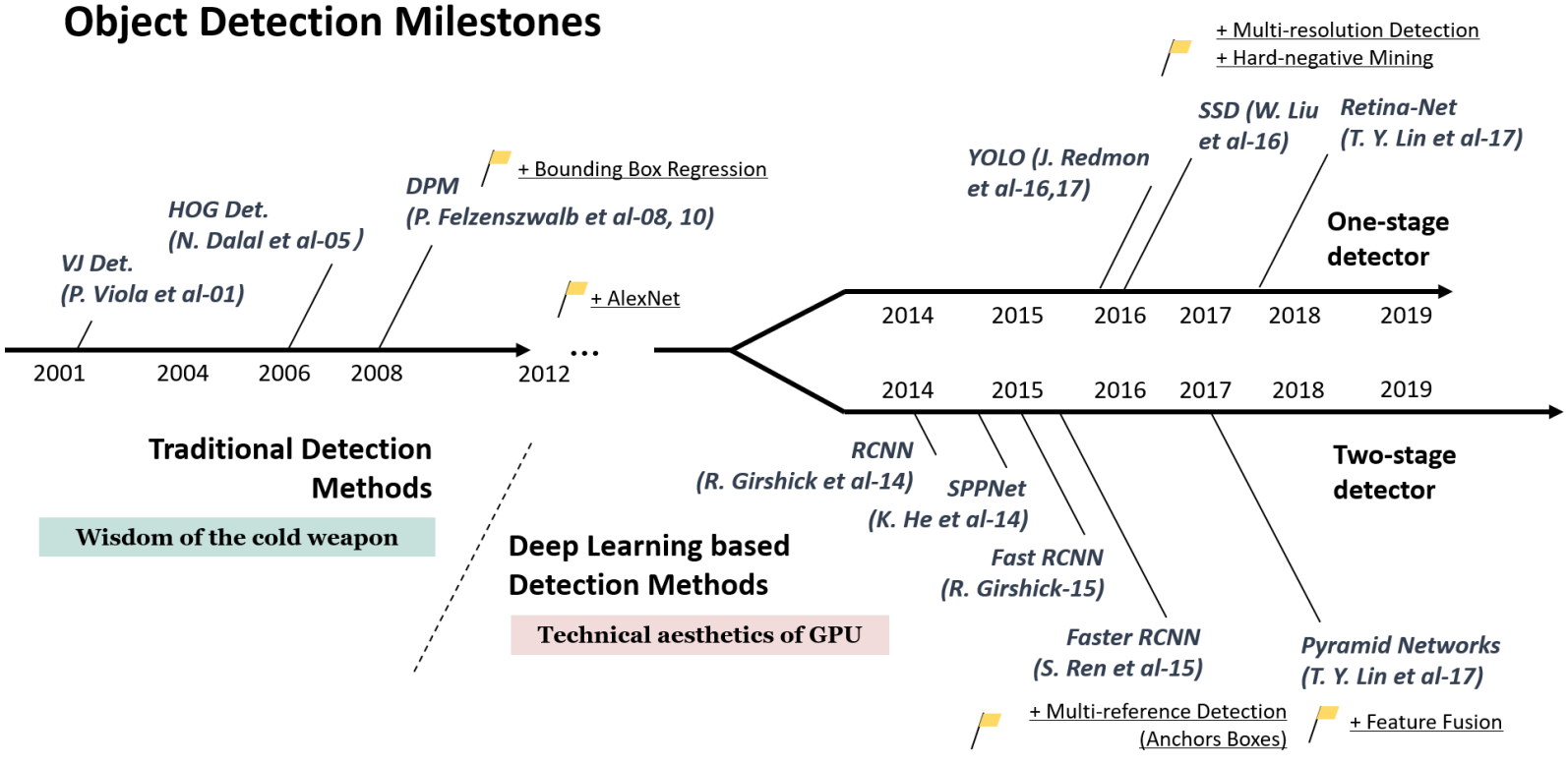
- 最重要的贡献是这个Loss
- 用这个Loss代替了交叉熵，极大提高了准确度
- 降低那些容易识别的类在Loss中的权重，提高那些难分类的
- : 准确预测概率

$$(\quad) = -(1 - \quad) \qquad (\quad)$$

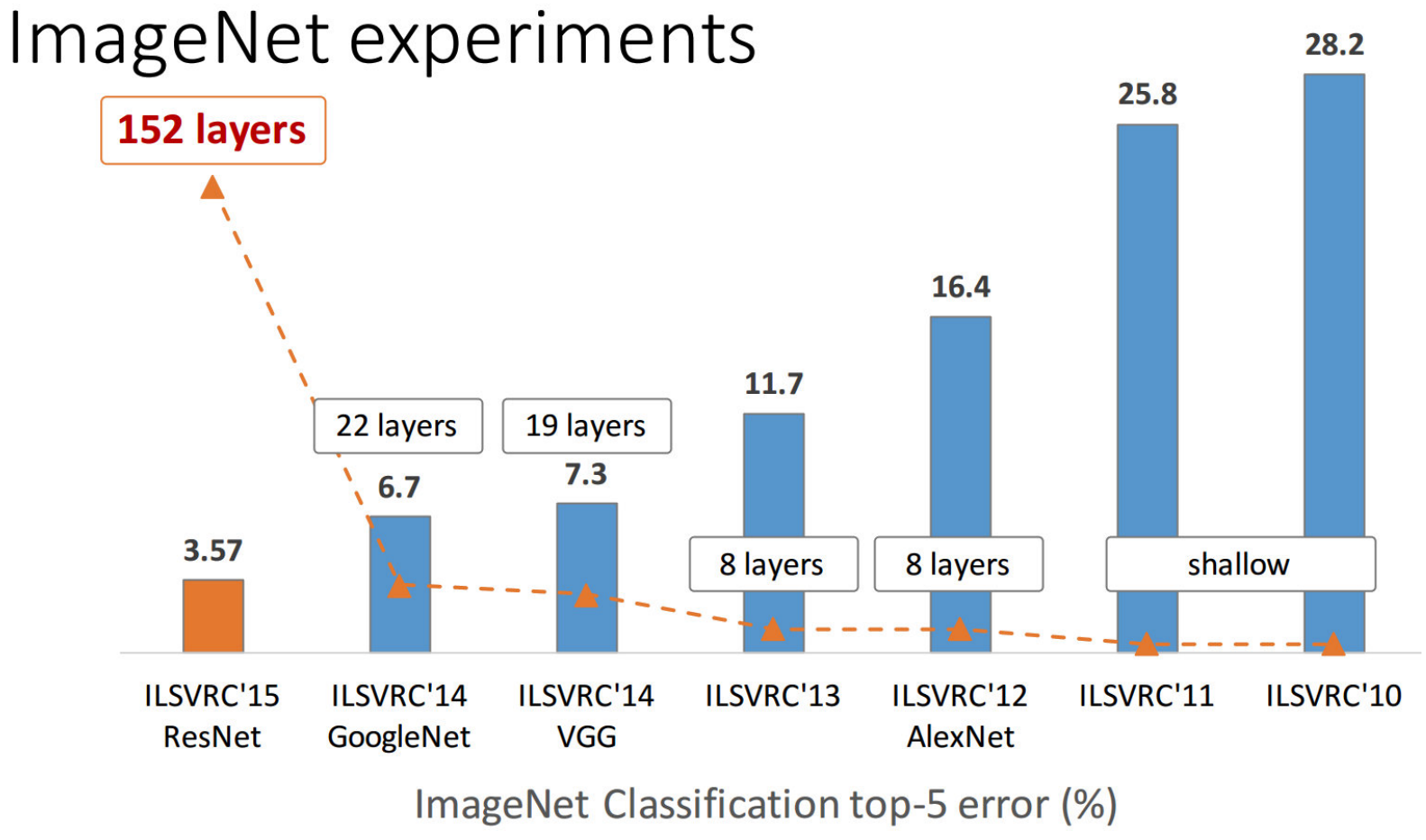
RetinaNet效果

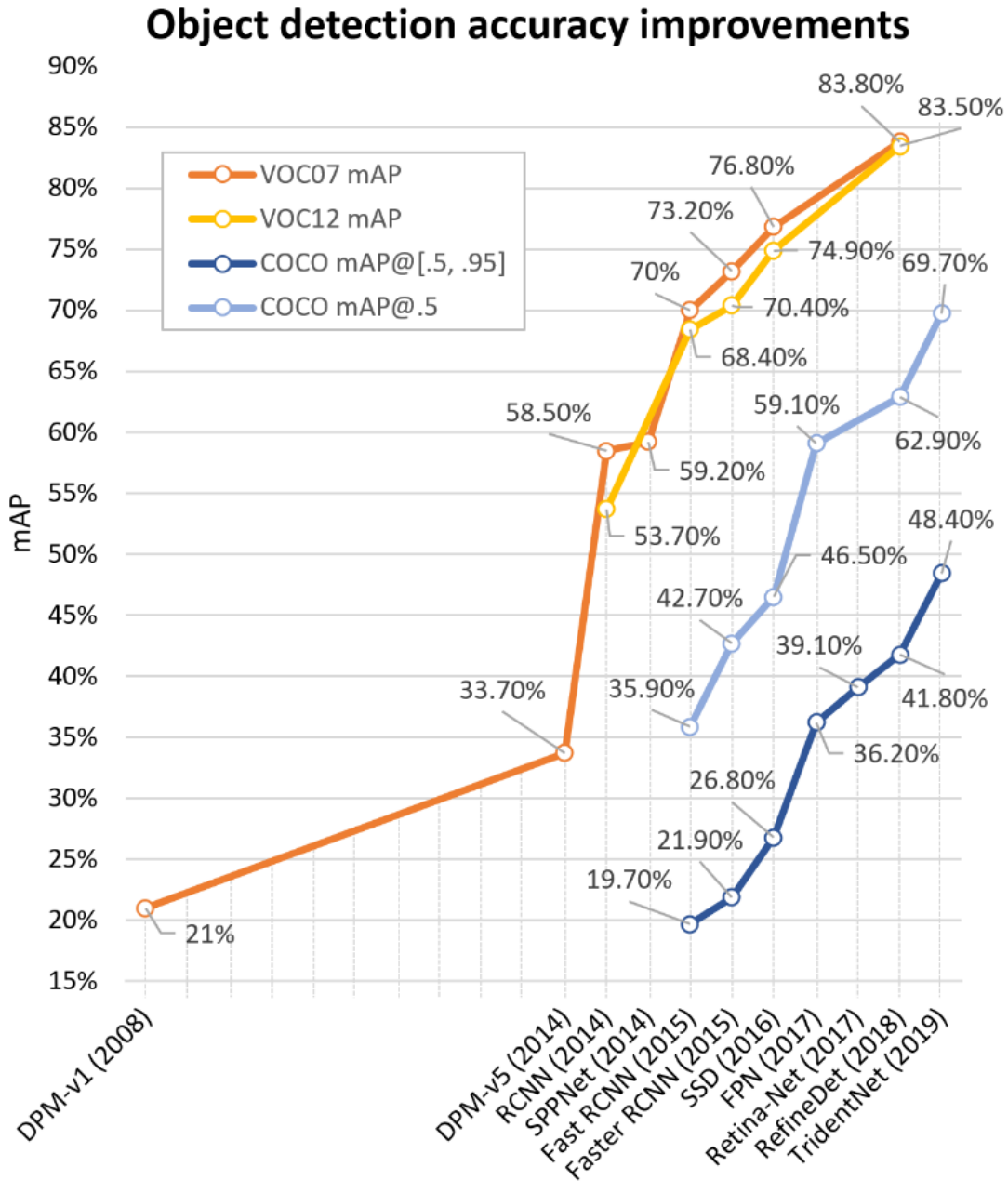


小结


Object Detection Milestones
+ Multi-resolution Detection
+ Hard-negative Mining
SSD (W. Liu et al-16)
Retina-Net (T. Y. Lin et al-17)
YOLO (J. Redmon et al-16,17)
One-stage detector
+ Bounding Box Regression
DPM (P. Felzenszwalb et al-08, 10)
HOG Det. (N. Dalal et al-05)
VJ Det. (P. Viola et al-01)
+ AlexNet
2001
2004
2006
2008
2012
2014
2015
2016
2017
2018
2019
2014
2015
2016
2017
2018
2019
RCNN (R. Girshick et al-14)
SPPNet (K. He et al-14)
Fast RCNN (R. Girshick-15)
Faster RCNN (S. Ren et al-15)
Pyramid Networks (T. Y. Lin et al-17)
Two-stage detector
Traditional Detection Methods
Wisdom of the cold weapon
Deep Learning based Detection Methods
Technical aesthetics of GPU
+ Multi-reference Detection (Anchors Boxes)
+ Feature Fusion

性能


ImageNet experiments
152 layers
22 layers
19 layers
8 layers
8 layers
shallow
3.57
6.7
7.3
11.7
16.4
25.8
28.2
ILSVRC'15 ResNet
ILSVRC'14 GoogleNet
ILSVRC'14 VGG
ILSVRC'13
ILSVRC'12 AlexNet
ILSVRC'11
ILSVRC'10
ImageNet Classification top-5 error (%)

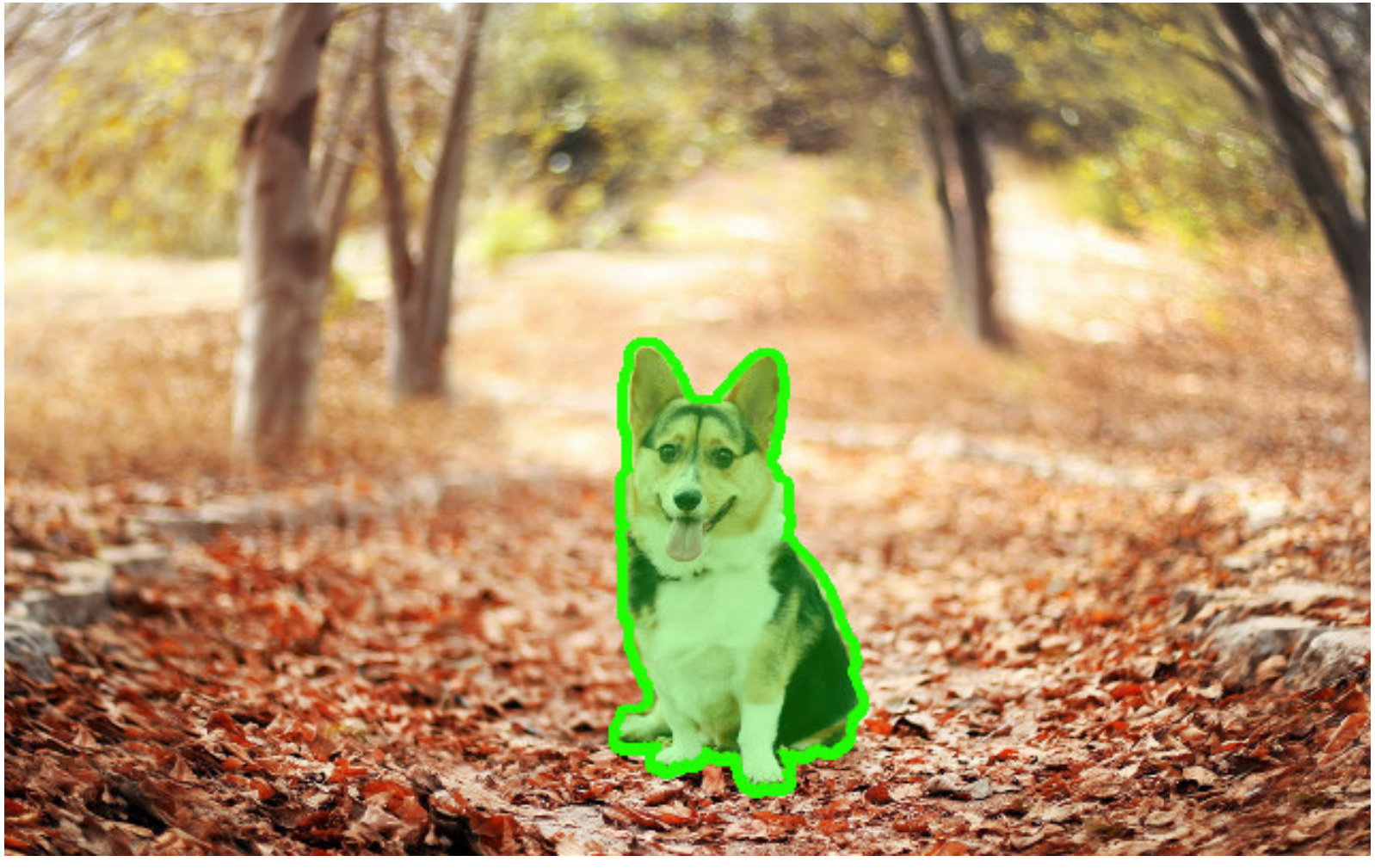

Object detection accuracy improvements
mAP
90%
85%
80%
75%
70%
65%
60%
55%
50%
45%
40%
35%
30%
25%
20%
15%
VOC07 mAP
VOC12 mAP
COCO mAP@[.5, .95]
COCO mAP@.5
21%
33.70%
53.70%
58.50%
59.20%
68.40%
70%
70.40%
73.20%
74.90%
76.80%
83.50%
83.80%
19.70%
21.90%
26.80%
35.90%
36.20%
39.10%
41.80%
42.70%
46.50%
48.40%
59.10%
62.90%
69.70%
DPM-v1 (2008)
DPM-v5 (2014)
RCNN (2014)
SPPNet (2014)
Fast RCNN (2015)
Faster RCNN (2015)
SSD (2016)
FPN (2017)
Retina-Net (2017)
RefineDet (2018)
TridentNet (2019)

图像分割

Object Segmentation

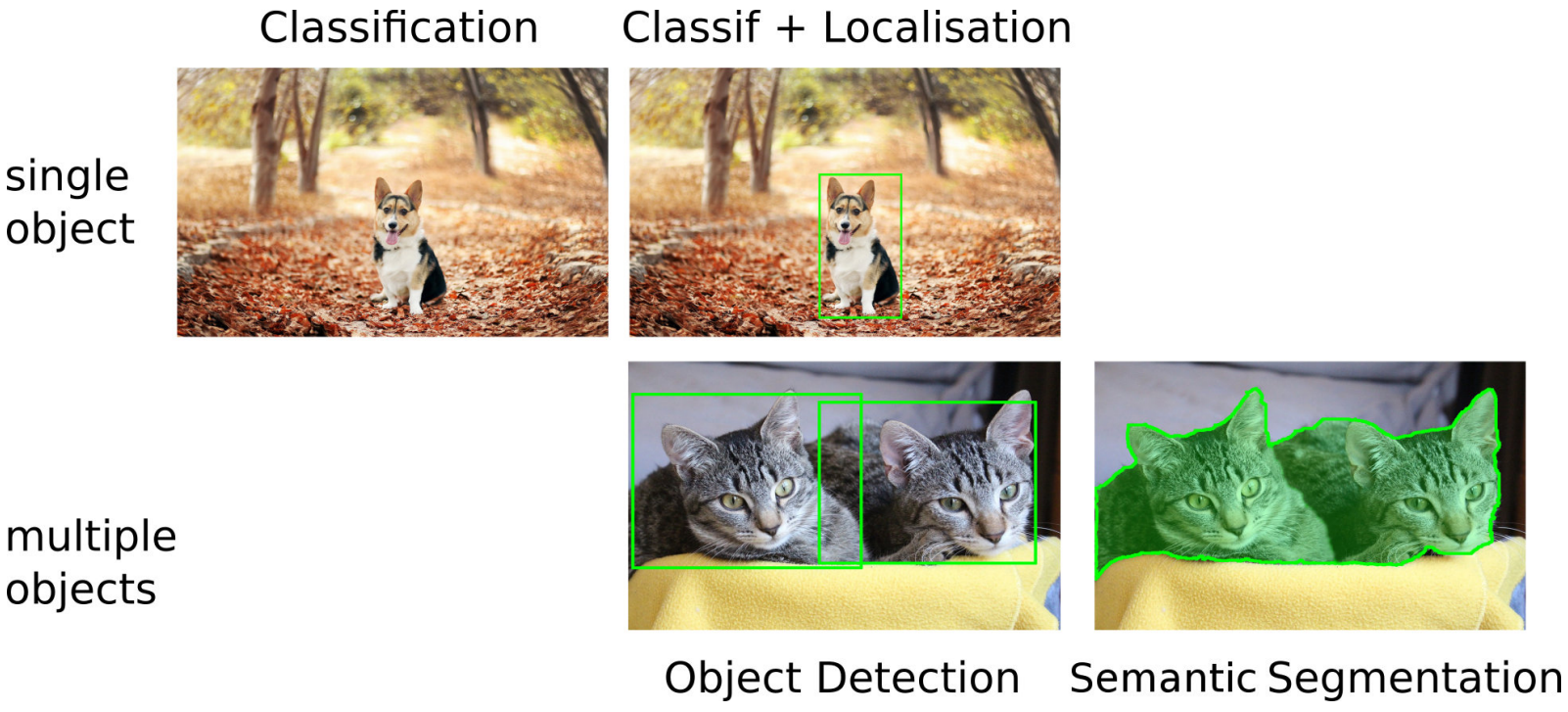
从图形中提取对象的轮廓

图像分割



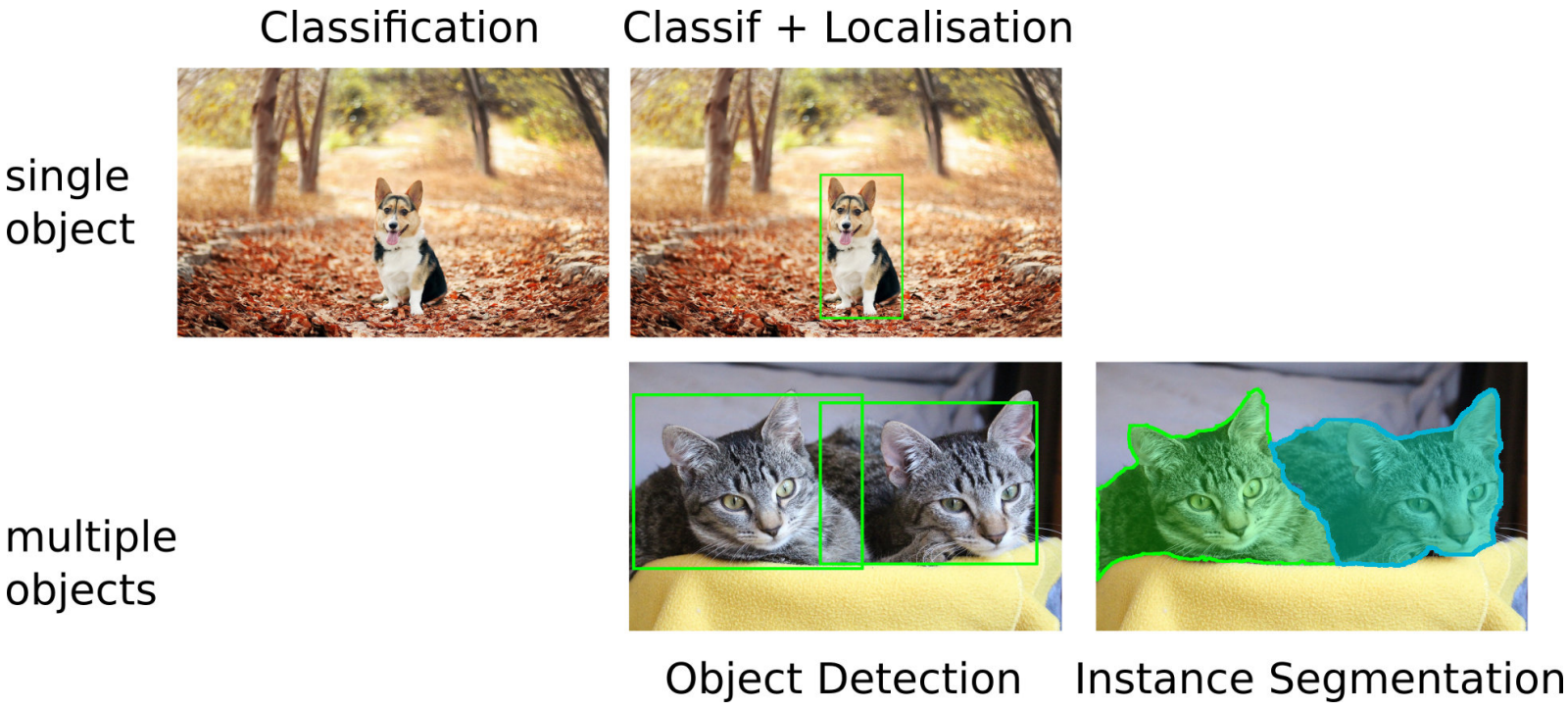
语义分割

Semantic Segmentation



实例分割

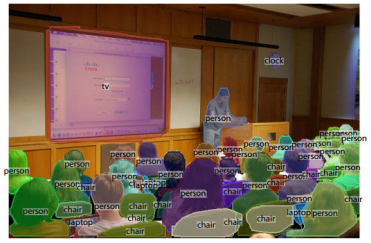
Instance Segmentation

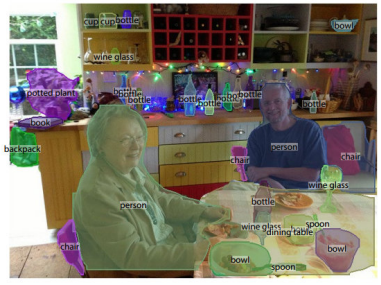


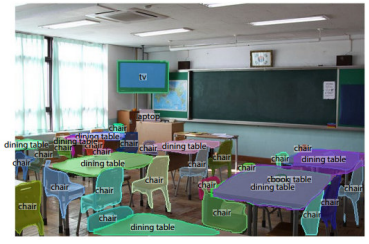
实例分割

实例分割





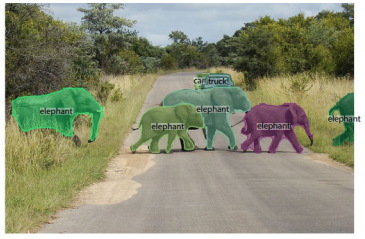


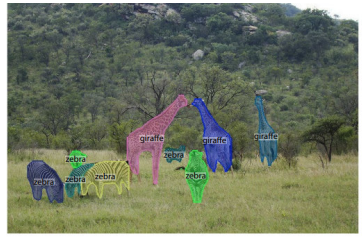


实例分割


sheep


zebra


elephant
car
truck


giraffe
zebra


bicycle
bus
umbrella
person


car
person
motorcycle
bus
truck


bus
person
truck


traffic light
person
backpack
handbag


person
suitcase
backpack


teddy bear
person


umbrella
handbag
person

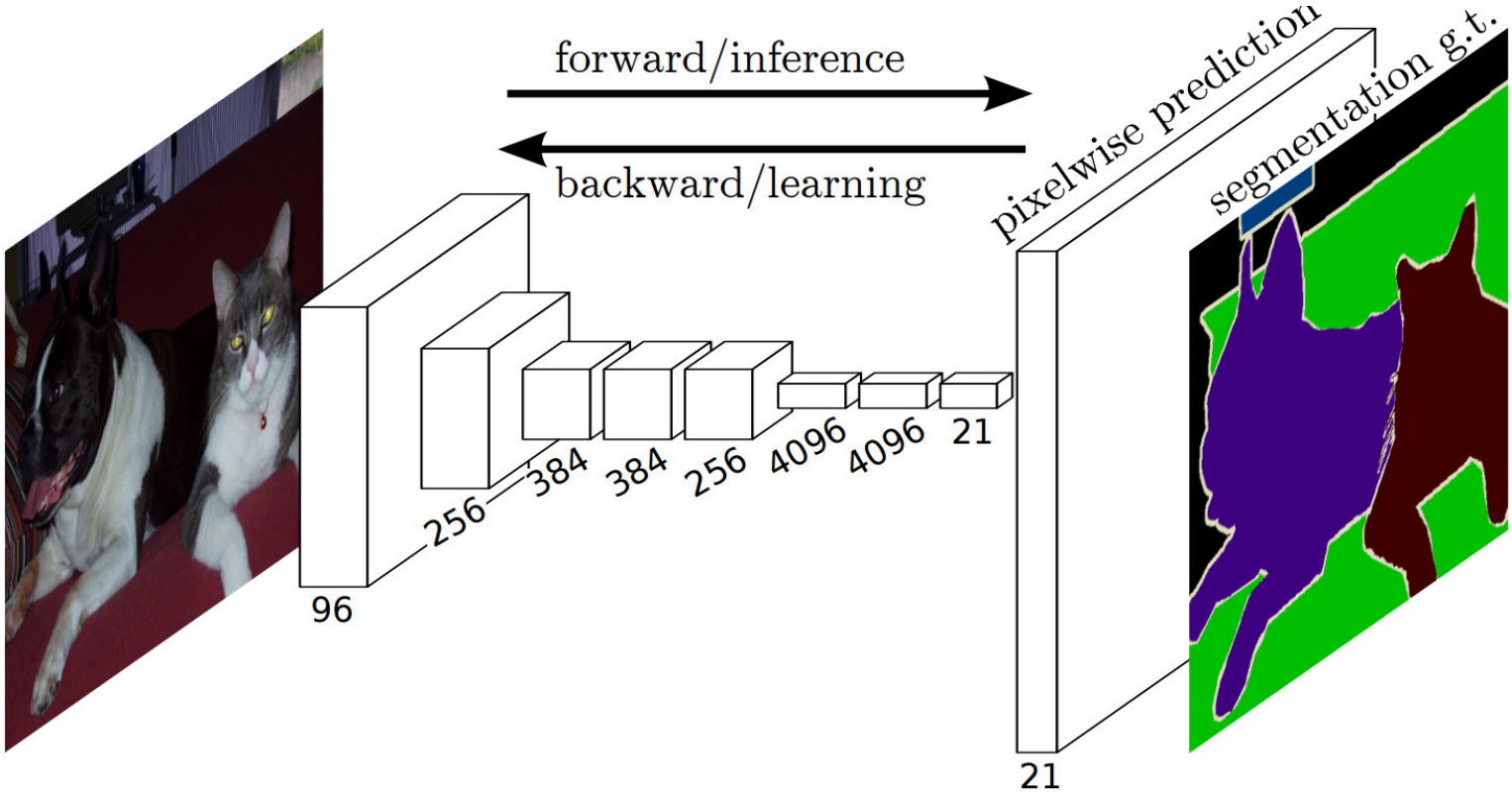

potted plant
vase

语义分割（2017）

00:00

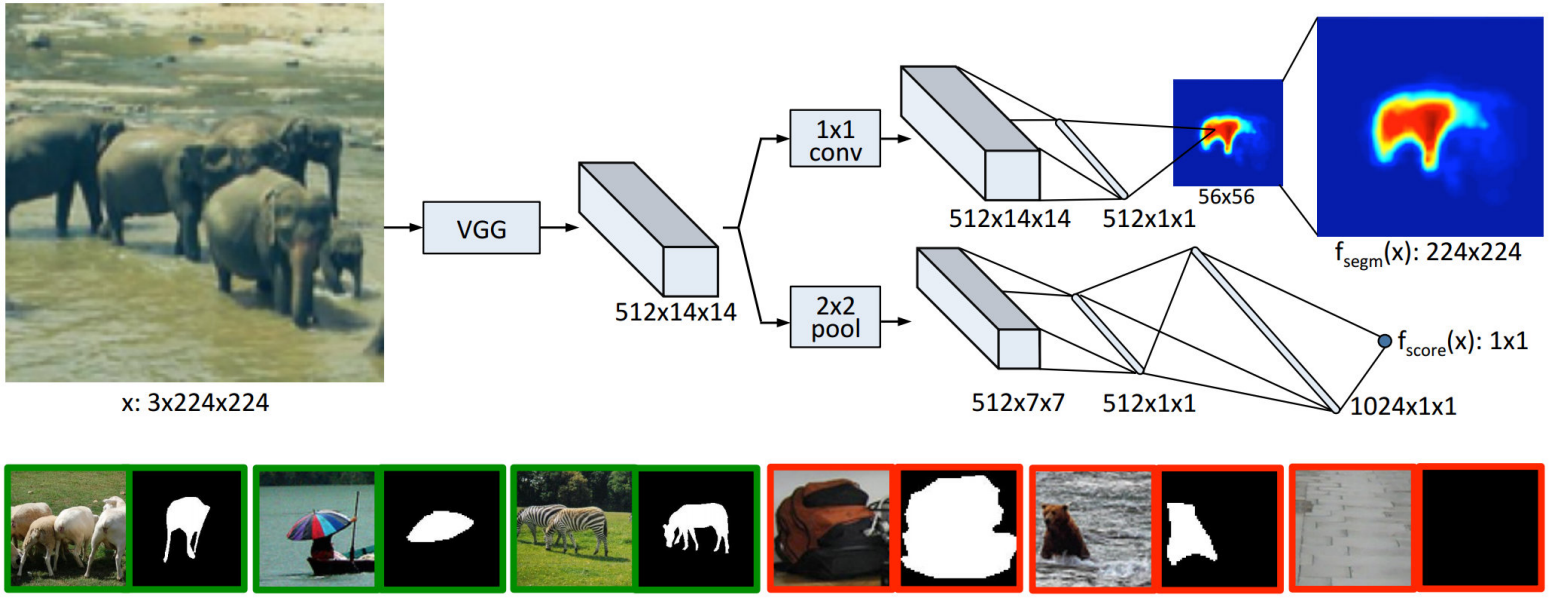
实例分割

分类每个像素，得到Mask



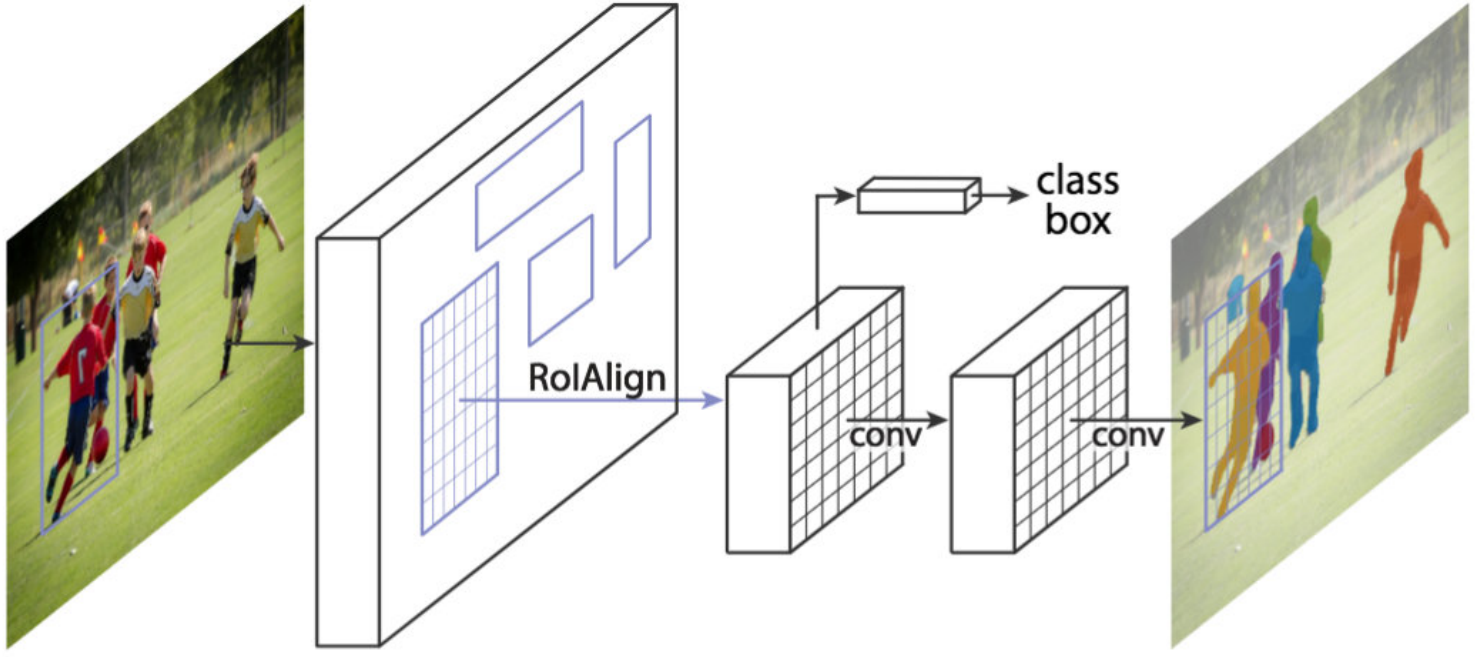
DeepMask

- Facebook，2015 NIPS
- VGG后分两路
  - MASK
  - 目标检测



Mask RCNN

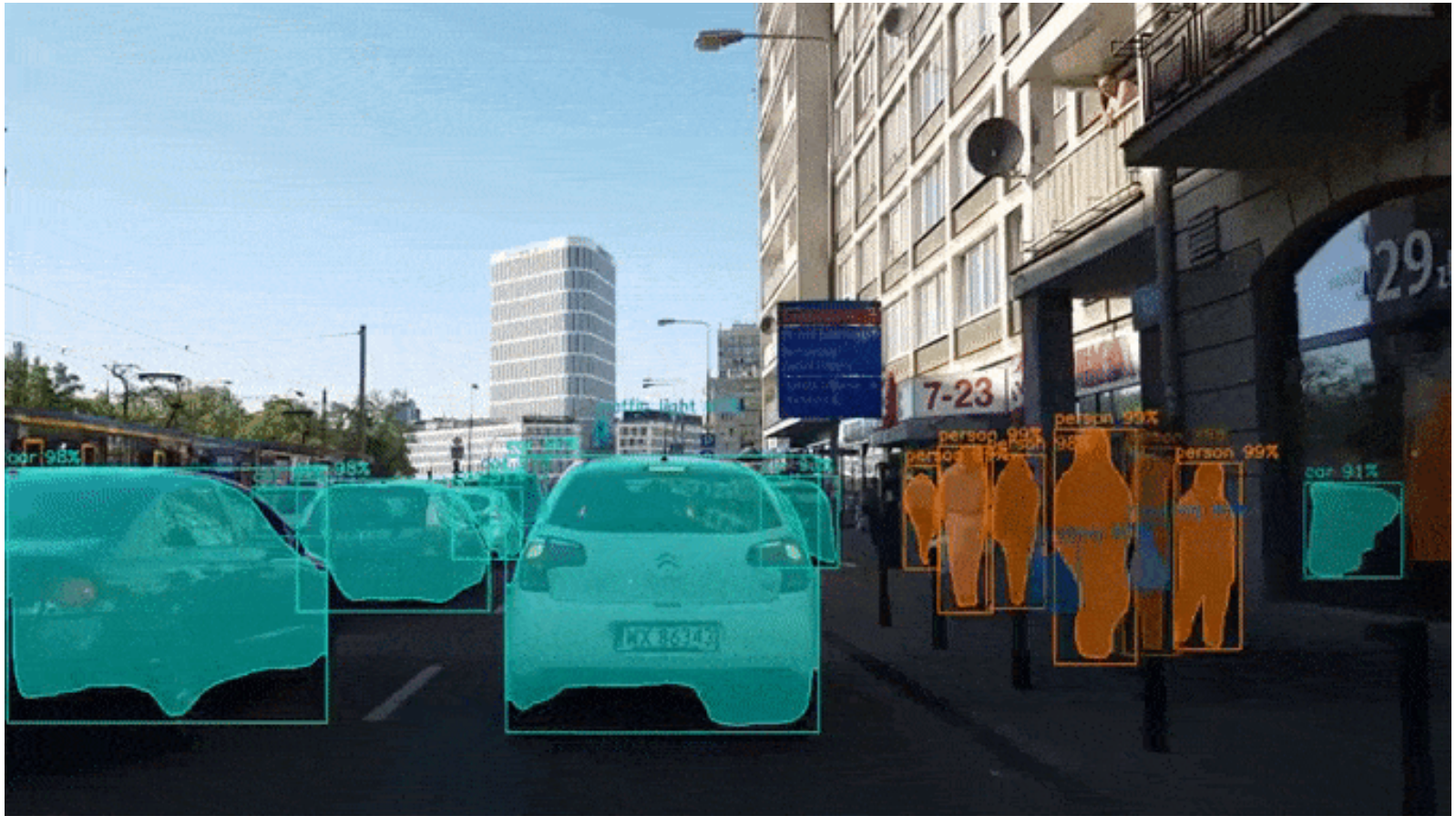
- 2017年，基于FPN（金字塔网络）和ResNet



Mask RCNN效果


person1.00
surfboard1.00
horse1.00
skateboard.82
kite.98
kite.97
kite.99
kite.82
kite.84
kite.93
handbag.80
skateboard.99
person.91
skateboard.98
car.87
car.93
car.95
car.97
car.99
car.78
traffic light.73
car.98
truck.86
bus1.00
car.1.00
car.82

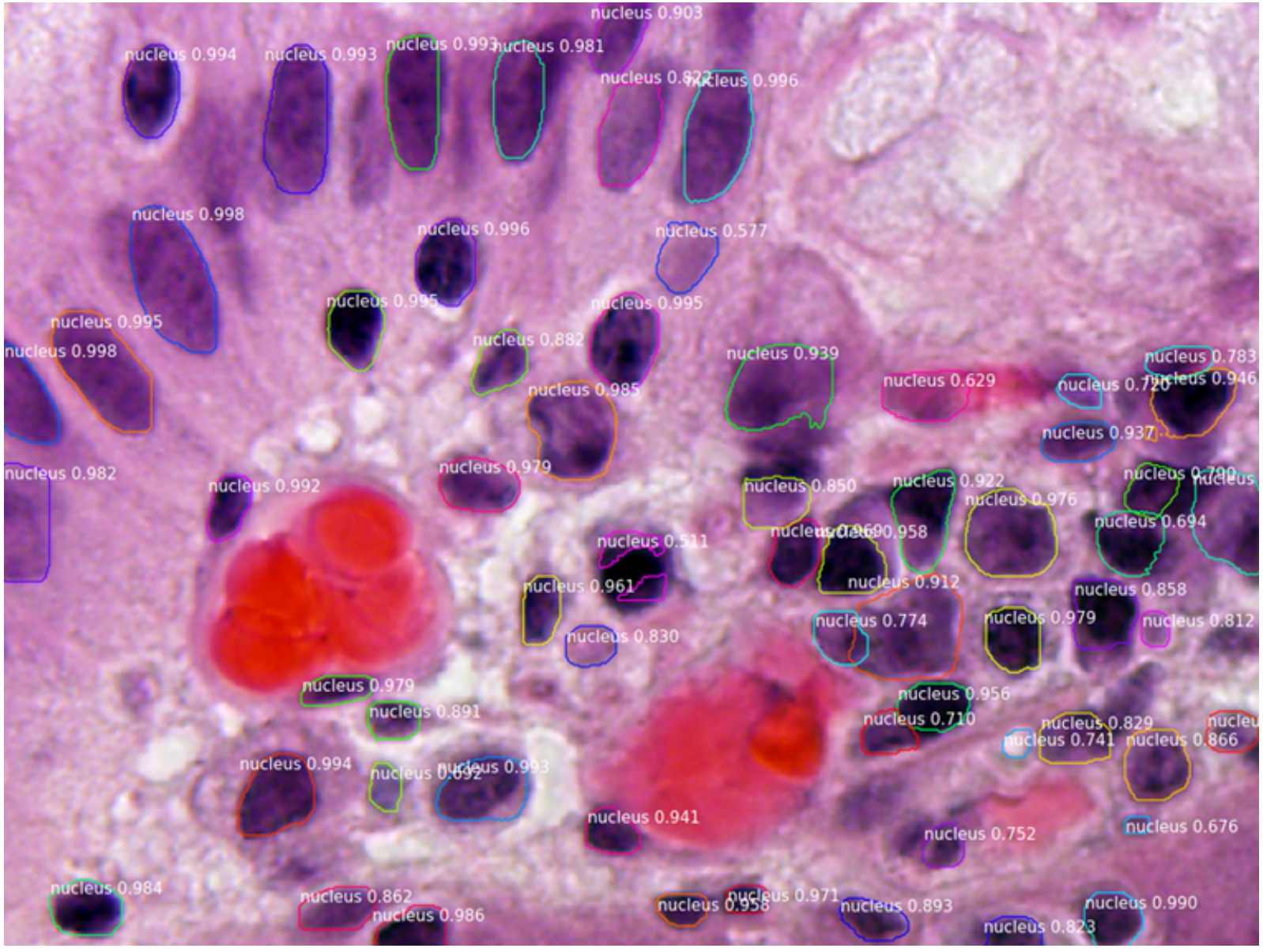
Mask RCNN效果



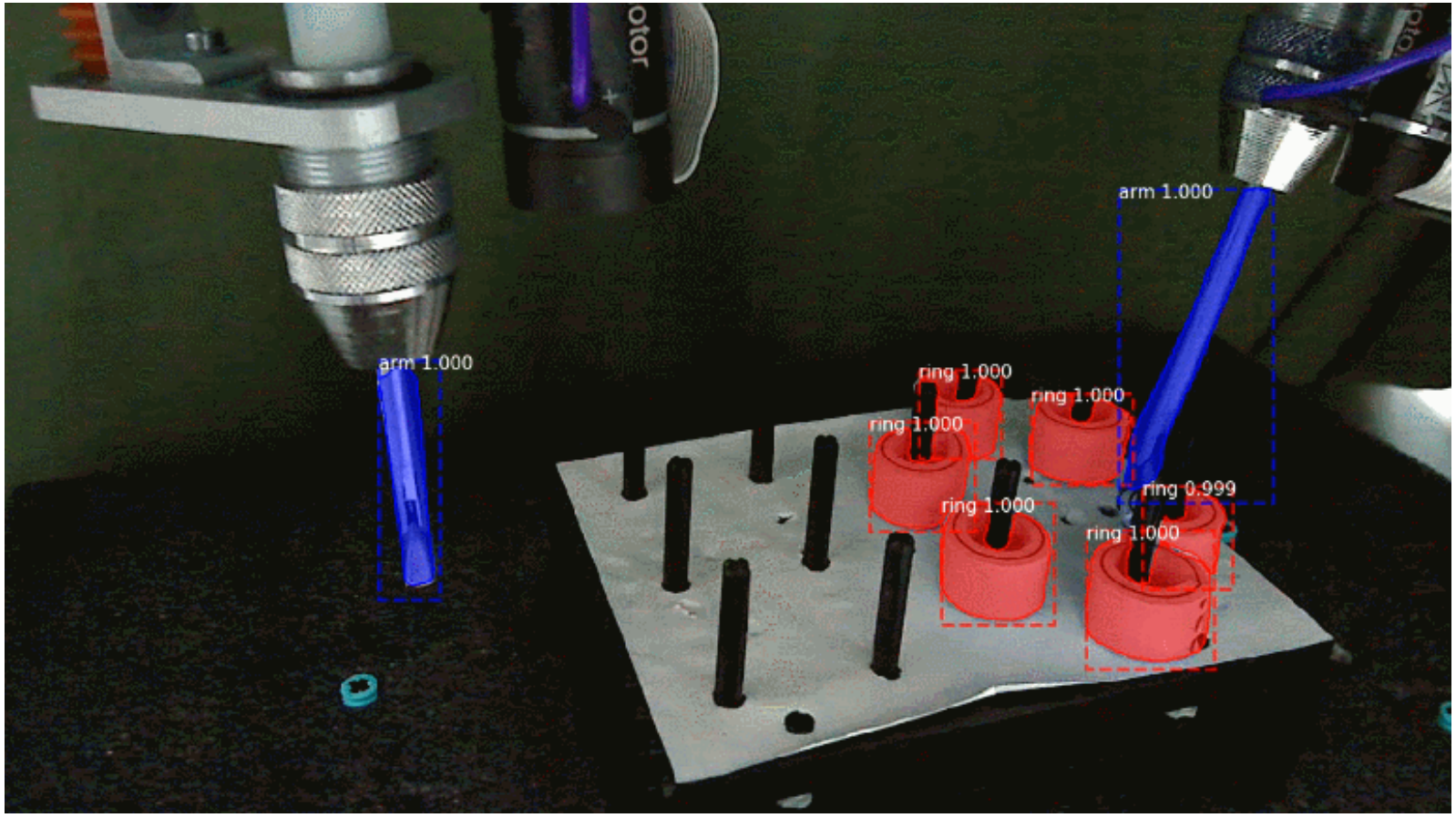
应用

颜色气球追踪

细胞核分割


nucleus 0.903
nucleus 0.994
nucleus 0.993
nucleus 0.993
nucleus 0.981
nucleus 0.996
nucleus 0.998
nucleus 0.996
nucleus 0.577
nucleus 0.995
nucleus 0.995
nucleus 0.995
nucleus 0.998
nucleus 0.882
nucleus 0.939
nucleus 0.783
nucleus 0.629
nucleus 0.720
nucleus 0.946
nucleus 0.985
nucleus 0.937
nucleus 0.982
nucleus 0.979
nucleus 0.992
nucleus 0.850
nucleus 0.922
nucleus 0.976
nucleus 0.694
nucleus 0.511
nucleus 0.961
nucleus 0.912
nucleus 0.858
nucleus 0.774
nucleus 0.979
nucleus 0.812
nucleus 0.830
nucleus 0.979
nucleus 0.956
nucleus 0.891
nucleus 0.710
nucleus 0.829
nucleus 0.741
nucleus 0.866
nucleus 0.994
nucleus 0.692
nucleus 0.941
nucleus 0.752
nucleus 0.676
nucleus 0.984
nucleus 0.862
nucleus 0.986
nucleus 0.971
nucleus 0.893
nucleus 0.990
nucleus 0.823

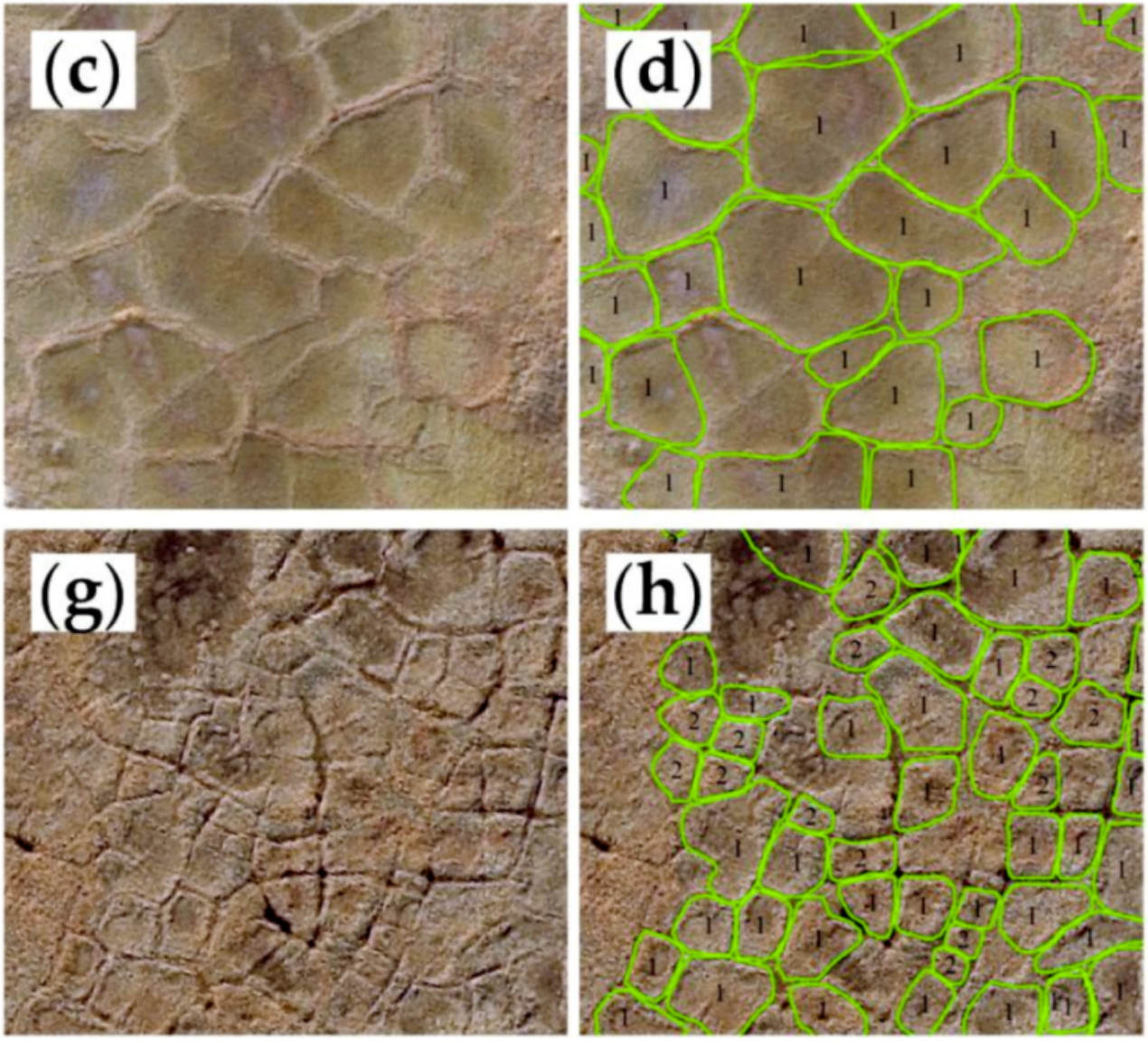
工业机器人


arm 1.000
arm 1.000
ring 1.000
ring 1.000
ring 1.000
ring 1.000
ring 0.999
ring 1.000

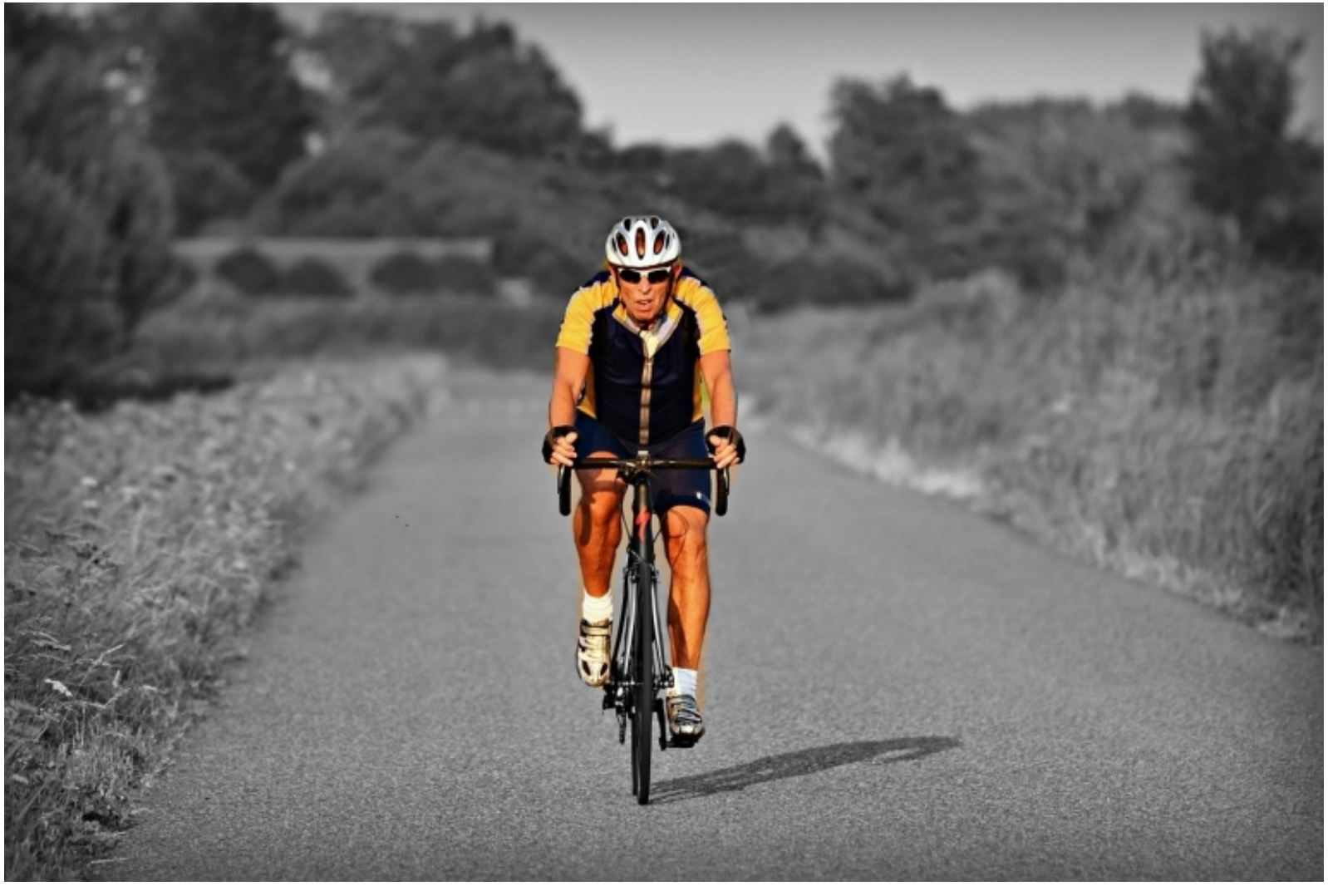
3D建筑物

细胞游动

地理多边形

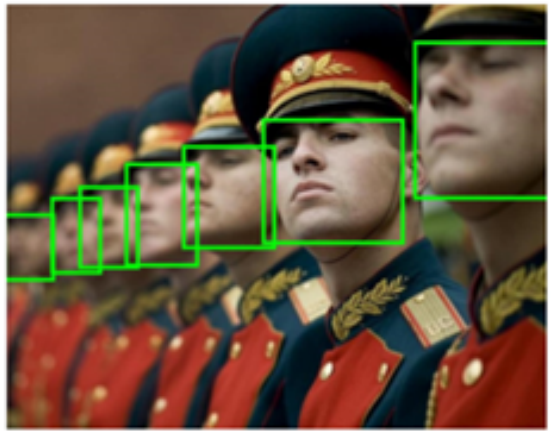


照片特效



人脸检测

(a)



(b)

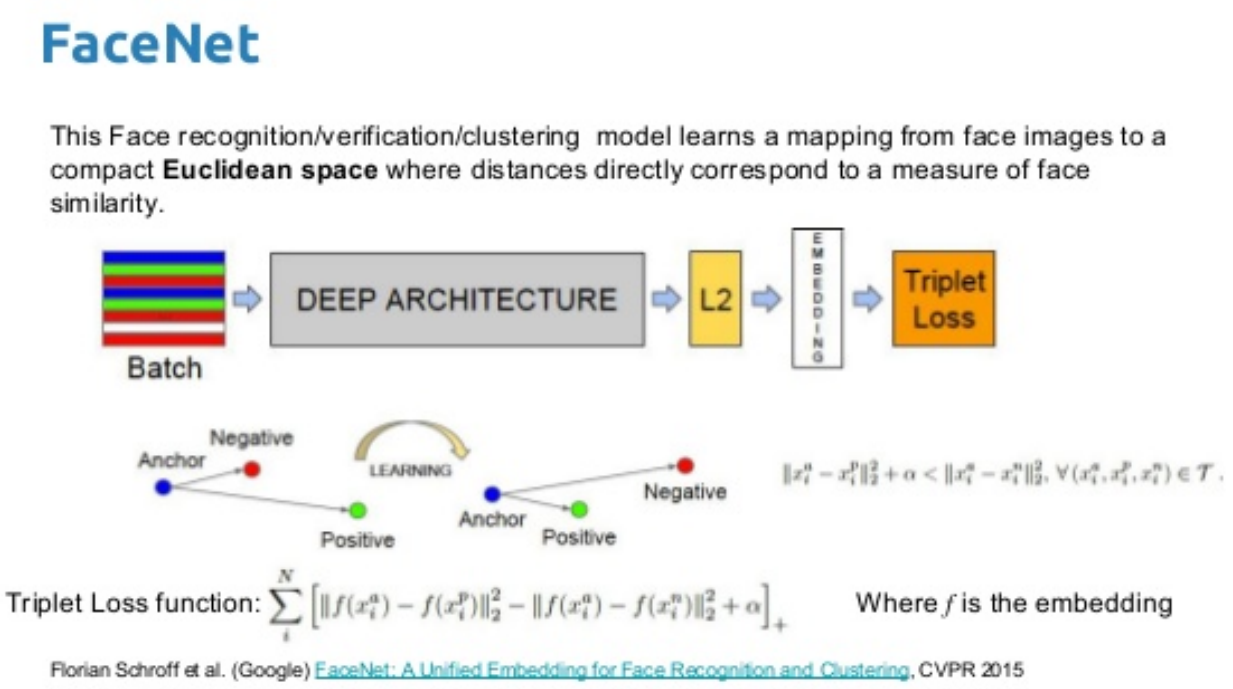
(c)

人脸识别

- 美国马里兰州枪击事件，人脸识别技术找出了嫌犯
- 流行歌星泰勒·斯威夫特，演唱会上过滤狂热粉丝和跟踪狂
- 收容所追踪收容所和避难所的使用情况

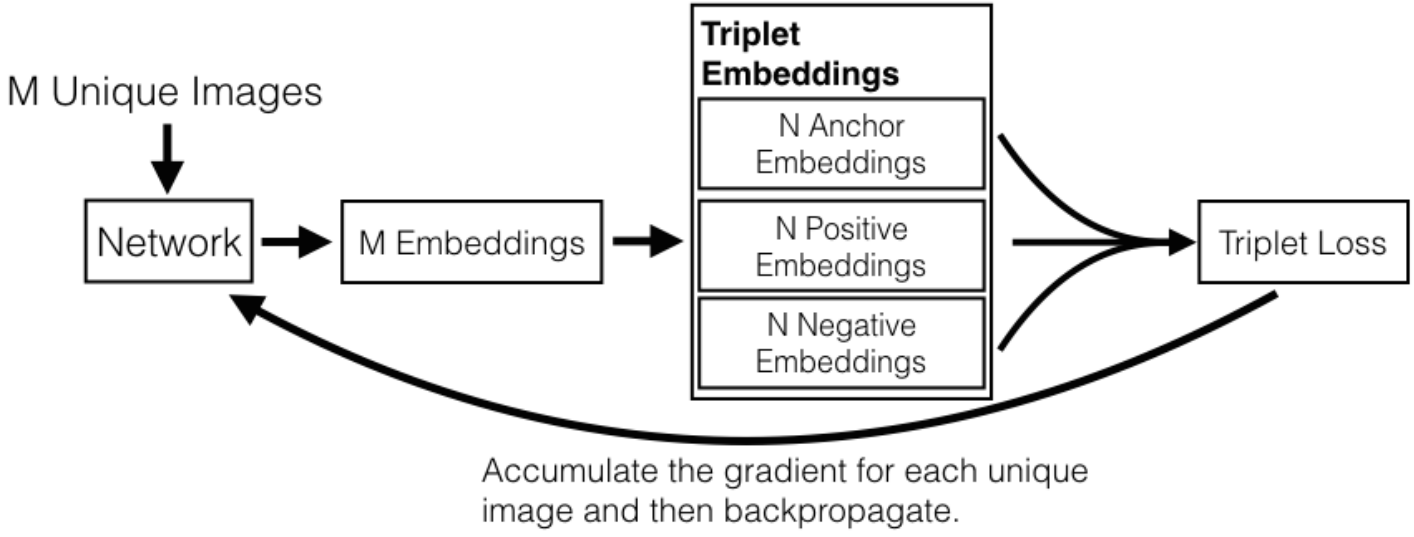
FaceNet

2015年Google提出



FaceNet架构

利用Triplet Loss捕获不同脸之间的相似和不同



FaceNet设计

将人脸转换为128维的向量表征

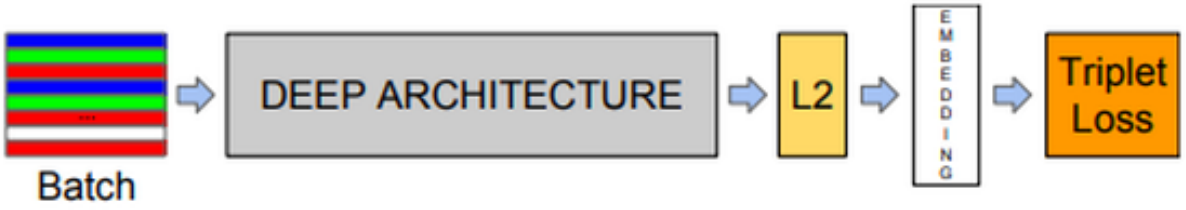


Figure 2. **Model structure.** Our network consists of a batch input layer and a deep CNN followed by L_2 normalization, which results in the face embedding. This is followed by the triplet loss during training.

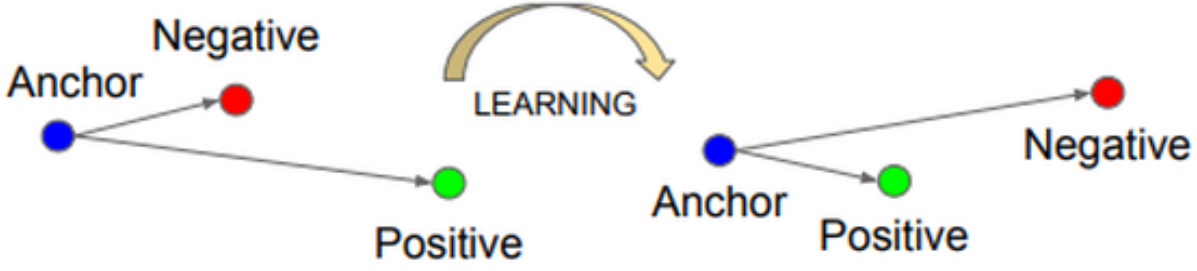


Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

姿势检测与识别

姿势检测与识别

情感


Engagement
100%
Brow Raise
5%
Brow Furrow
0%
Valence
100%
Lip depressor
0%
Smile
100%

交通流量计数

386 人在观看： 「木JJ出品」2018第18周 Billboard美国单曲… 立即围观 >

人工智能：交通流量计数

去bilibili观看 分享

播放器初始化...[完成]
加载用户配置...[完成]
加载视频地址...[完成]
加载视频内容...

00:00 / 00:00 360P

进入bilibili,一起发弹幕吐槽!

去吐槽

交通流量计数

441 人在观看： 2018 SEPHOR HAUL！全脸新品试用try on … 立即围观 >

人工智能：交通流量计数II

去bilibili观看 分享

播放器初始化...[完成]
加载用户配置...[完成]
加载视频地址...[完成]
加载视频内容...

00:00 / 00:00 360P

进入bilibili,一起发弹幕吐槽! 去吐槽

交通信号识别

(a)

(b)

(c)

铁轨检测

483 人在观看： 小学生：我是你爸！打一巴掌还得赔2000元！ 立即围观 >

人工智能：铁路信号检测

去bilibili观看 分享

播放器初始化...[完成]
加载用户配置...[完成]
加载视频地址...[完成]
加载视频内容...

00:00 / 00:00 360P

进入bilibili,一起发弹幕吐槽!

去吐槽

道口监控

836 人在观看： [310]战神全剧情娱乐流程解说01 立即围观 >

人工智能：铁路信号检测

去bilibili观看 分享

播放器初始化...[完成]
加载用户配置...[完成]
加载视频地址...[完成]
加载视频内容...

00:00 / 00:00 360P

进入bilibili,一起发弹幕吐槽!

去吐槽

应用：文本识别


龍城大華
DRAGON CITY
海龍賓館
HOI LOON HOTEL
HOTEL FACILITEIS OVER NIGHT LESS 20%
REGENG HOTEL
moca hair
We love hair
麗晶賓館
REGENG HOTEL
FOOT
TEL:2388 209
Venus
沙嘴王
龍中西
館賓晶

(a)


CITYLINK

(b)


£100 CASHBACK
THIS WEEKEND ONLY
ON ACER ASPIRE LAPTOPS

(c)

存在的问题

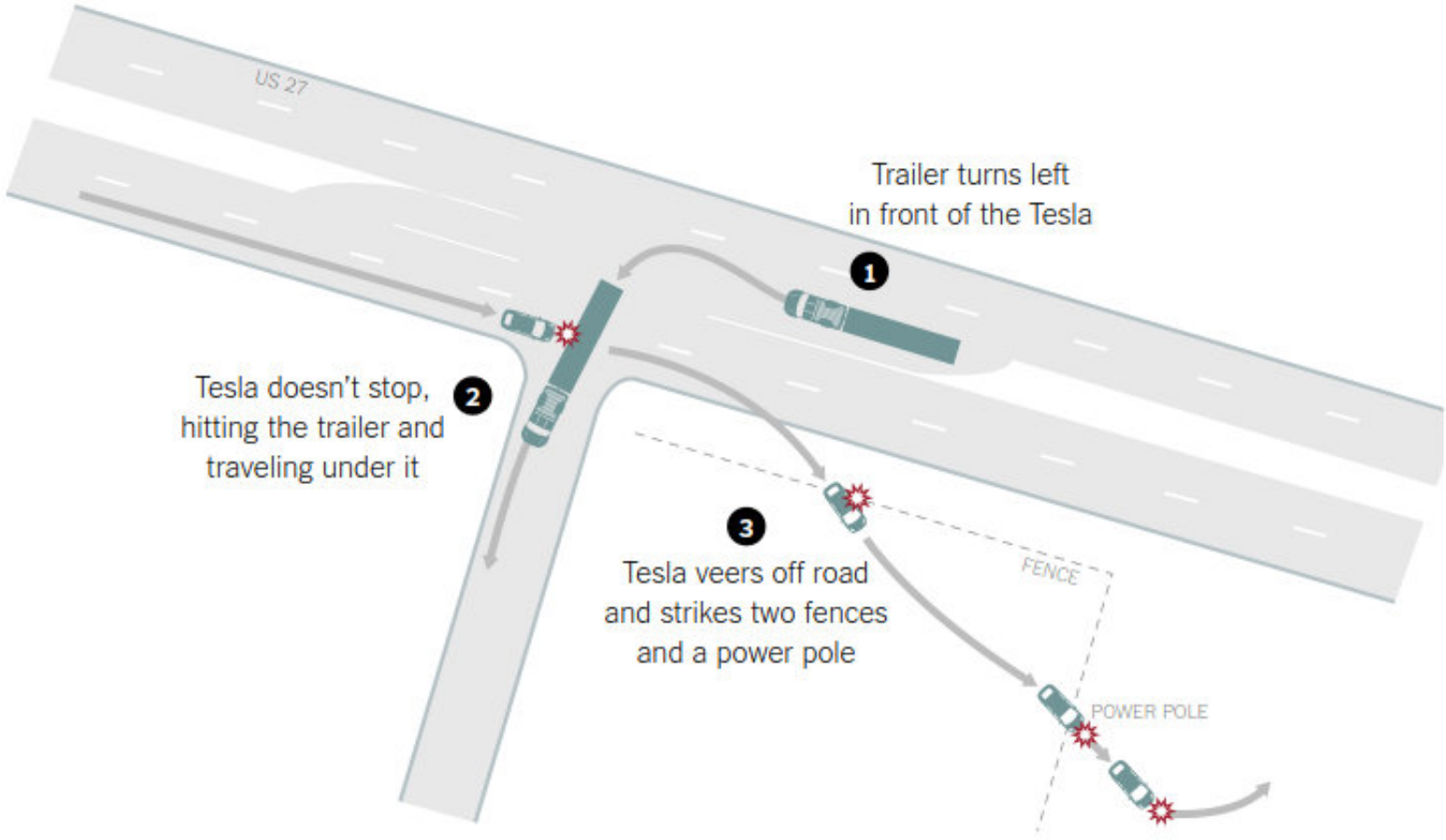
准确率、隐私保护、公平

准确率问题

- 2018年7月，“美国公民自由联盟”测试了亚马逊的人脸识别系统
- 将535名国会议员面孔，对照25000张公开的警方嫌疑犯照片。有28个无辜的国会议员被认成了嫌疑犯
- 对于皮肤较黑的人和女性，人脸识别通常不太准确。所有国会议员的错误率是5.2%，非白人国会议员错误率达39%

识别错误付出生命代价

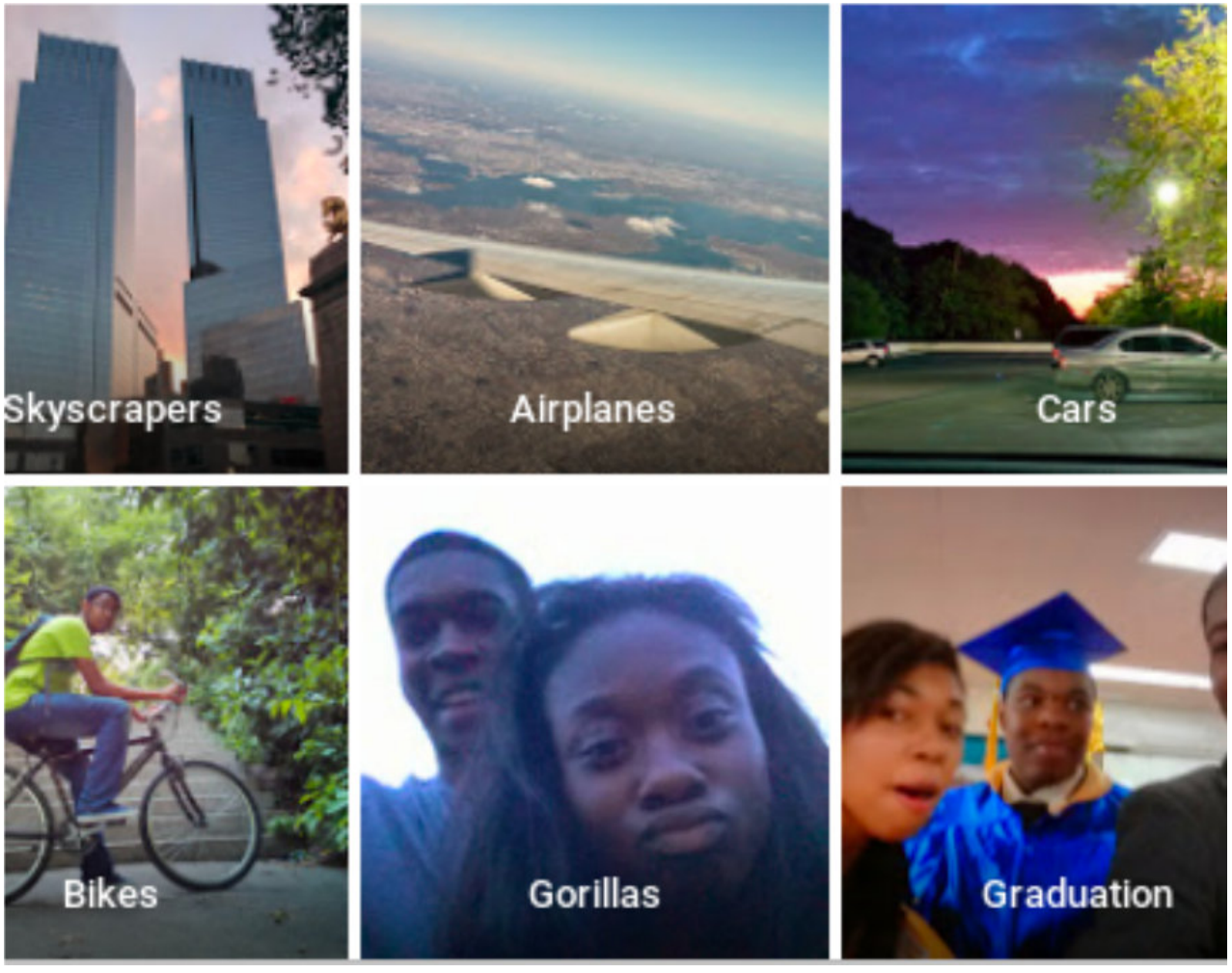
特斯拉自动驾驶系统未成功识别出白色货车



识别错误付出生命代价

识别错误引起民愤

把人识别成大猩猩



隐私保护

- 2019年5月14日，旧金山城市监督委员会以8票对1票通过法令，禁止城市工作人员购买和使用人脸识别技术
- “人脸识别技术危害公民权利和公民自由的倾向大大超过了其声称的好处，这项技术将加剧种族不平等，并威胁到我们不受政府长期监控的生活能力”

小测验

- 实例分割是做什么用的?
- 现实中，计算机视觉技术应用需要注意哪些问题?
- 举例说明你工作中可能需要的计算机视觉应用
- 深度学习在图像领域带来重大突破，请举出一个令你印象深刻的例子
- HOG算法通过计算图像的方向梯度直方图能够得到图像的什么特征?