



# 车辆检测与型号识别项目 总结报告

彭湃 原卉  
Peng Pai, Yuan Hui

2018.09.09



**车辆检测及型号识别** 广泛应用于物业，交通等的管理场景中。通过在停车场出入口，路口，高速卡口等位置采集的图片数据，对车辆的数量、型号等进行识别，可以以较高的效率对车型，数量信息等进行采集，并可以在不同的场景中辅助不同业务的开展，如商场停车位的规划，路况规划，或者公安系统追踪肇事车辆等等。

# 技术路线

## 任务理解

Task Understanding



### 目标

车辆检测与型号识别

+

### 数据

车辆数据集  
无bbox  
有细分类标签

公开数据集  
(PASCAL\_VOC,  
COCO中车辆图片)  
含有bbox  
无细分类标签

寻找方案

## 解决思路

Multiple Solutions



### 强监督学习

检测模型

分类模型

模型融合

+

### 弱监督学习

特征激活图 (CAM)

利用分类所依据特征的位置  
对目标进行定位

落实方案

## 方案实现

Plan Implementation



### 模型实现

#### 模型搭建

强监督模型  
弱监督模型

#### 模型训练

Learning rate  
Batch size  
Optimizer  
...

+

### 评价指标

accuracy  
预测准确度

IoU  
标注重合度

+

### 系统搭建

输入输出环境搭建



# 任务 | 01

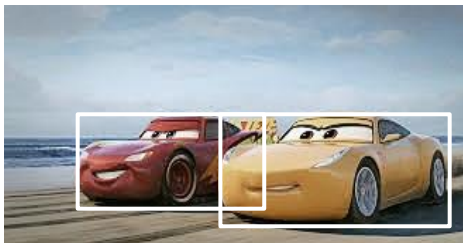
The Task

项目需求与数据准备

# 项目需求

## 车辆检测 + 型号识别

- 搭建一个系统，对用户上传的任意一张图片进行车辆检测，并显示出车辆位置信息
- 如果图片中检测到车辆，需对每一辆车进行型号识别，并显示出型号信息
- 如果图片中没有检测到车辆，可以给出没有检测到的提示



Class A



Class B

# 数据准备



val\_1500\_label\_  
669



val\_1600\_label\_  
582



val\_1700\_label\_  
199



val\_1800\_label\_  
515



val\_2400\_label\_  
351



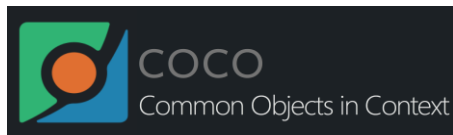
val\_2500\_label\_  
595



val\_2600\_label\_  
49



val\_2700\_label\_  
105



## 项目提供的数据集

- 共 48856 张图片， 764 个分类。其中 43971 张作为训练集， 4885 张作为验证集。
- 只有 image level 的标注， 没有 instance level 的标注
- 观察训练图片和验证图片， 每张图片有一个目标， 或明显有一个主要目标， 一个 label

## 已有的公开数据集

- 由于项目提供的数据集不含有位置信息， 可尝试利用一些包含车辆位置信息的公开数据集来辅助检测任务
- 包含 car/bus 的 PASCAL\_VOC 数据集： train 和 val 共 1441 张
- 包含 car/bus 的 COCO 数据集： train 14078 张， val 622 张



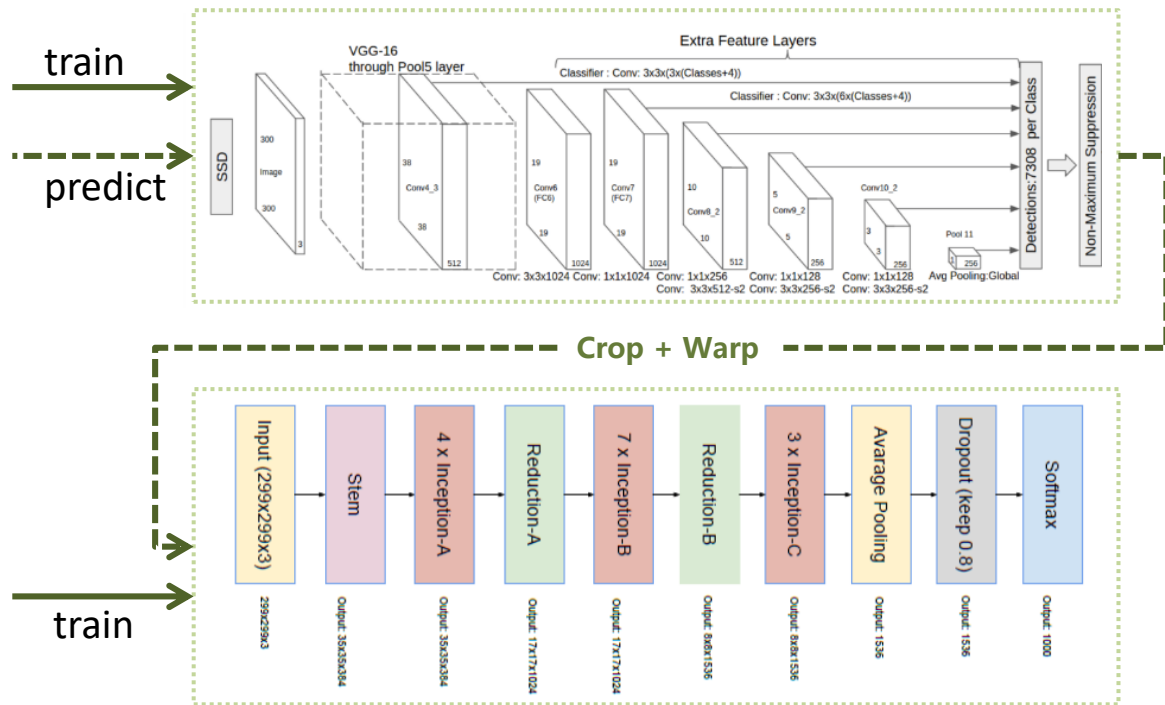
# 思路 | 02

Solutions

强监督方案 vs. 弱监督方案

# 强监督：网络设计

## SSD + Inception v4



### 模型构思:

- 可以把“检测”和“识别”任务拆分成 2 个模型来完成。
- 模型一：用于检测车辆位置信息。任务所提供的没有车辆位置信息，可借助包含车辆信息的公开数据集来搭建检测模型。
- 模型二：用于做车辆分类任务。车辆型号数据集已经提供。

### 模型训练:

- 两个模型分别训练。

### 预测:

- 先送入检测模型，根据检测出的 BBox 进行裁切后，送入分类模型。



## 强监督：实验结果

### 检测正确的结果示例



Label: 雪铁龙-富康



Label: 英菲尼迪-FX



Label: 双龙-享御

## 强监督：实验结果

### 检测错误的结果示例

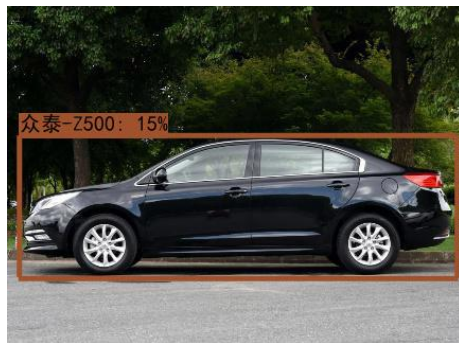
- 实验中发现，有两类图片识别的错误率较高：
  - 在训练集中出现次数较少的车型
  - 车辆侧面的图片
  - 车型本身比较相似



Label: 英菲尼迪-FX



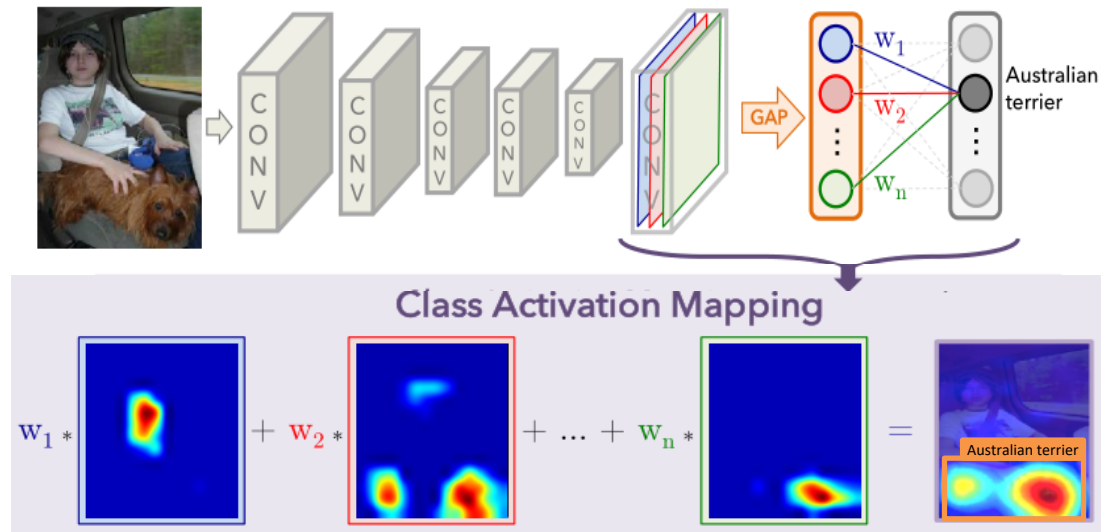
Label: 捷豹-XJ



Label: 众泰-Z700

## 弱监督：网络设计

### 理解CAM (Class Activation Map)



#### 基本思想：

- 图片在经过卷积层的特征提取后，还保留了一定的空间信息，可以协助对分类的依据进行定位。

#### CAM生成方式：

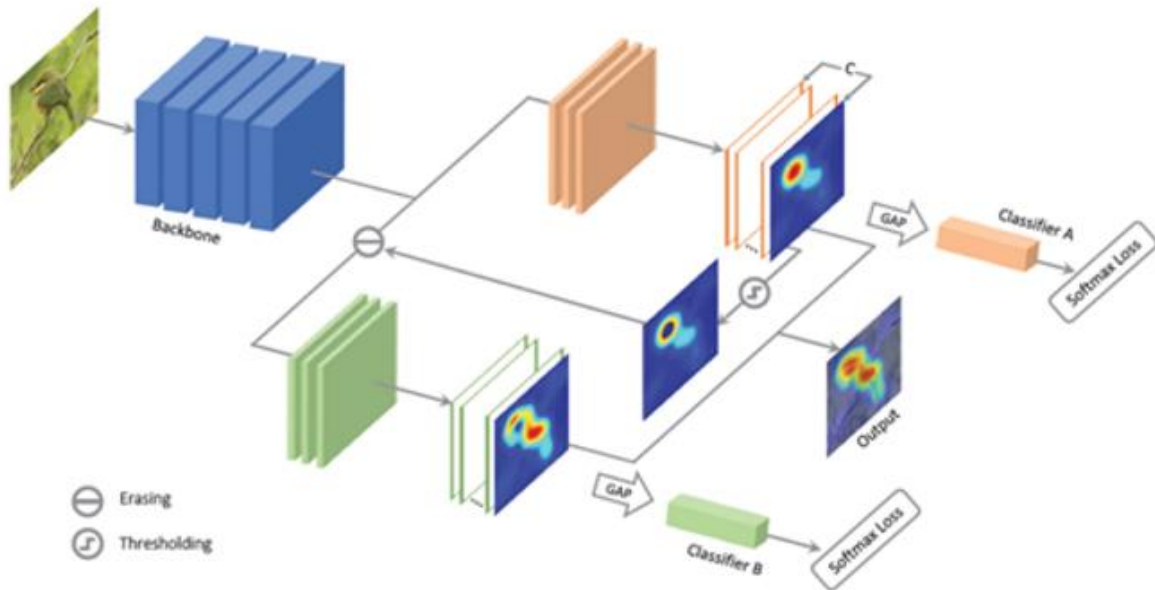
- 提取特征后，使用Global Average Pooling，再通过全连接送入分类层。将最后一层的特征图和相应类别全连接层权重相乘后可得到CAM。

#### bounding box划定：

- 设置阈值，并框出超过阈值的最大相连区域。

## 弱监督：网络设计

### 对CAM的改进：Adversarial Complementary Learning



#### 改进一：

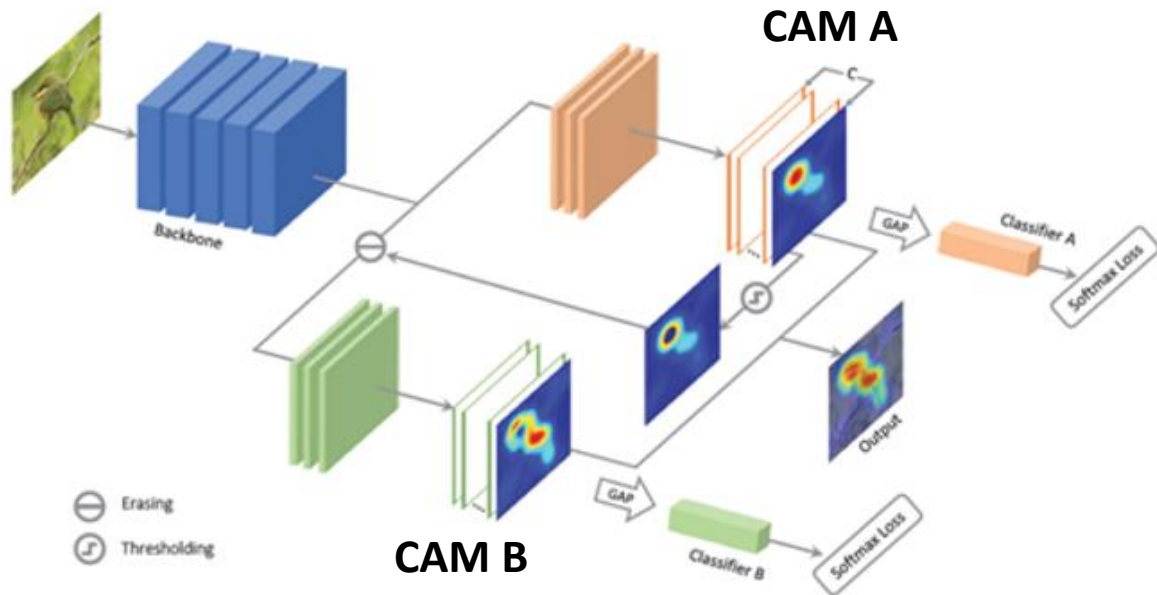
- 简化了CAM的生成方式，可以在网络训练过程中直接生成CAM
- 先根据类别数目进行特征整理，再进行 global average pooling生成特征向量进行分类。

#### 改进二：

- 采用两个分类器，生成互补的特征区域，实现更完整的物体定位。

## 弱监督：实现细节

### 基于inception\_v4网络的CAM实现



#### 基础网络:

- inception\_v4 (保留最后一个inception block 之前的结构)

#### Branch A:

- 加上两个分别有 1024 个  $3 \times 3$  的卷积核的卷积层 (采用same padding)
- 加上一个有类别数个  $1 \times 1$  的卷积核的卷积层
- 进行global average pooling

#### Branch B:

- 将A生成的CAM标准化到 $[0,1]$ 区间, 大于阈值 $\delta$ 的为A的特征区域 ( $\delta$ 设为0.8), 在 inception 提取的特征图中, 将这些区域置零
- 后面结构同 branch A

#### 生成CAM:

- 融合两个classifier 生成的heatmap (取每个像素位置的最大值)
- 采用双线性插值恢复到原图尺寸

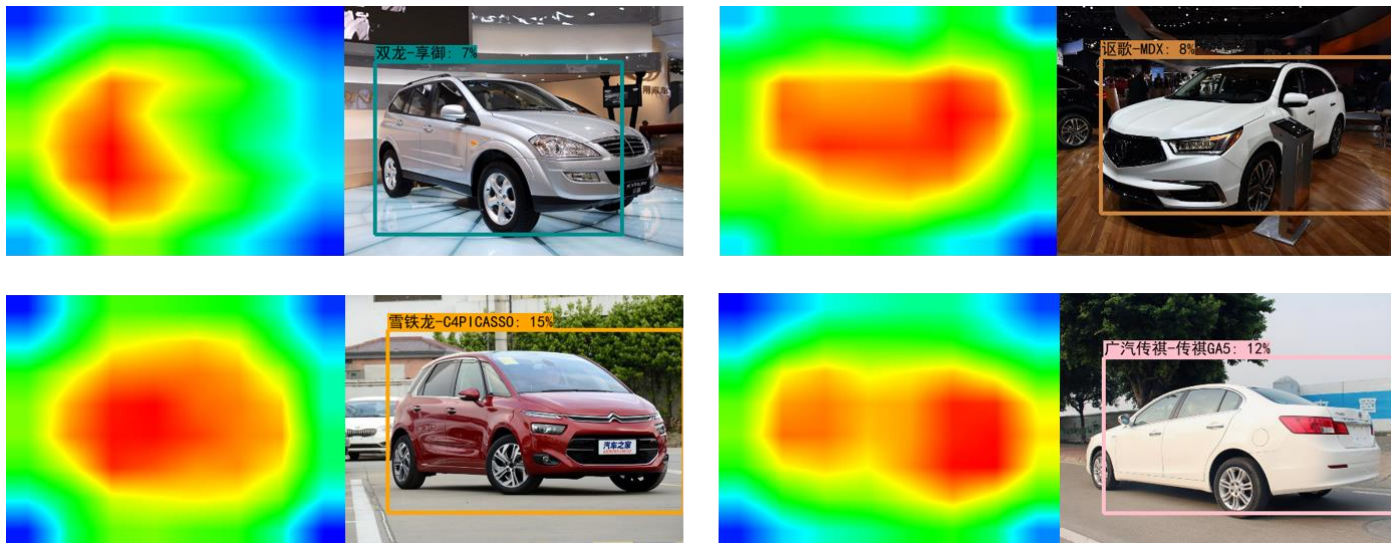
#### 划定bounding box:

- 阈值设为 0.75

## 弱监督：实验结果

### 分类正确：较好的检测结果示例

- 对于分类正确的物体，CAM有能力给出相对合理的定位结果

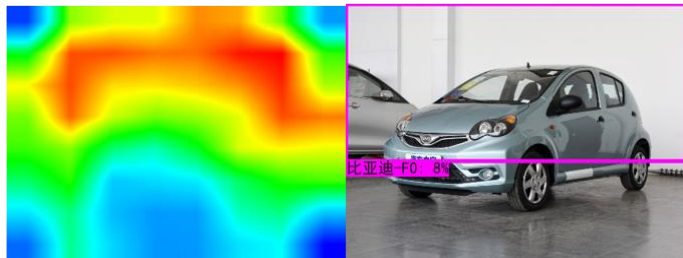
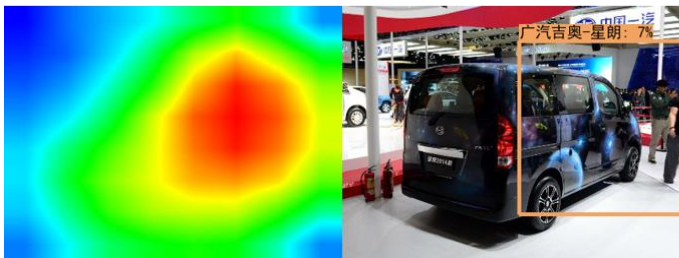
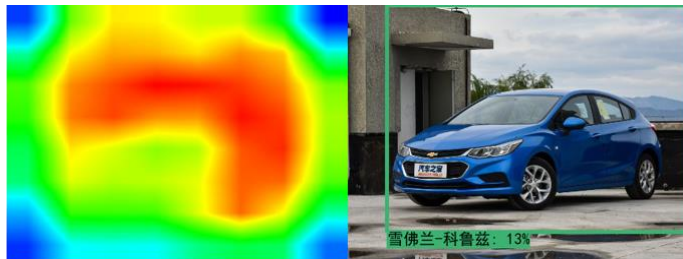
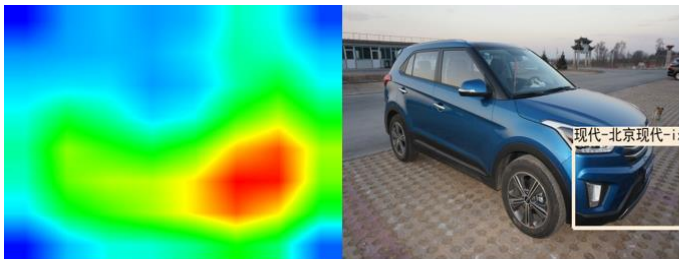




## 弱监督：实验结果

### 分类正确：较差的检测结果示例

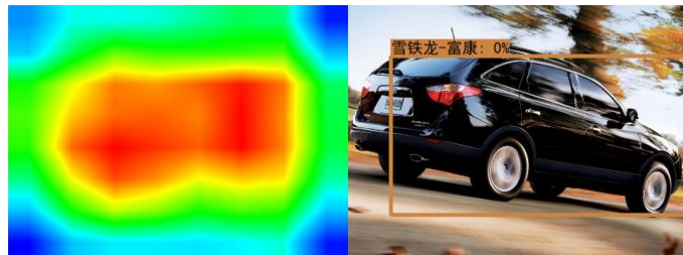
- 得分较高的特征区域不能完全覆盖车辆的全部，主要原因在于，并非物体的所有部分都具有区分性（比如车窗、轮胎等），即不能都对分类产生较大的贡献
- 得分较高的特征区域较车辆的实际位置有明显的偏移，可能是 CAM 的上采样过程所造成的偏差



## 弱监督：实验结果

### 分类错误的结果示例

- 对于分类错误的物体，其分类结果基本上无法指导定位，但个别情况下，也会产生一些巧合性的结果。其原因可能是有些车型本身或图片所展示的方向辨识度低（或与其他车型相似），仅提取出了车辆的某些共同特征（或相似特征），仍可对定位起到一定指导作用。





## 弱监督：存在问题

### 不易实现多目标检测

- 以分类为基础，主要解决定位任务，对于多目标的检测任务较难实现。
- 可以输出多个分类的结果（比如 top5），但要指定类别数，且对于多个同类物体不适用

### 上采样过程会增大偏差

- 采用 inception\_v4 作为基础网络时，最终生成的 CAM 尺寸为 8x8，而原图尺寸在 300-400 左右，还原时放大倍数在 30 倍以上，会造成较大偏差
- 未来可考虑的改进方向：调整网络结构，利用浅层网络保留的位置信息辅助进行逐步升采样，同时保留较准确的定位和较好的分类能力

实现

Implementation

03

模型评价、性能提升与演示环境搭建

# 模型评价

## 评价指标

- 由于网络以分类任务为基础，可将验证集上的分类准确率作为分类性能的评价指标
- 由于验证集中没有给出位置信息，我们在验证集中手动标出了 200 张图片的 bounding box，在分类正确的图片中，通过 IoU 的平均值评价预测框和人工标注框的重合程度

## 方案比较

- 强监督方案: Accuracy 0.883, Recall\_5 0.960, Mean IoU 0.867
- 弱监督方案: Accuracy 0.822, Recall\_5 0.952, Mean IoU 0.598
- 总体上说，弱监督方法的局限性还比较大，在效果上还是无法与强监督学习相比的

# 性能提升

## 性能提升目标

- 深入研究强监督方向，提升系统性能，让系统更接近商业化水准

## 从训练数据集上优化

- 数据增强：从某网站上抓取的数据，共 250083 张图片，200 个品牌，1836 个车系（车型）
- 图片预处理：
  - 用训练出的检测模型，把所有图片标记上BBOX（策略：confidence>80%，BBOX最大，h/w>0.4，w/h>0.6）
  - 裁切出BBOX内的区域，送入分类模型进行训练（经实验证明可以提高准确率）
- 实验结果：最终分类 accuracy 达到 0.922，Recall\_5 达到 0.989
  - 在新的训练集和验证集上有更高的准确率
  - 可以支持更多车型的识别

# 系统搭建

## Web演示地址

- <https://bdim.vvpen.com/prod/#/demo2>



上传图片



检测结果



识别结果



Thank you