**《用户手册》**

**图像分类系统**

**项 目 名 称： 图像分类系统**

**成 员 名 单： 王迪 马鲜艳 范广宝 陈晨**

**导 师： 张曙**

**工 程 领 域： 图像分类**

**研 究 方 向： 物体的分类与检测**

**中国科学技术大学软件学院**

目录

[目录 2](#_Toc12966843)

[一. 引言 3](#_Toc12966844)

[1.1 编写目的 3](#_Toc12966845)

[1.2 项目背景 3](#_Toc12966846)

[1.3 参考资料 3](#_Toc12966847)

[二. 软件概述 5](#_Toc12966848)

[2.1 目标 5](#_Toc12966849)

[2.2 功能 5](#_Toc12966850)

[2.3 性能 6](#_Toc12966851)

[三. 运行环境 7](#_Toc12966852)

[3.1 硬件 7](#_Toc12966853)

[3.2 支持软件 7](#_Toc12966854)

[四. 使用说明 8](#_Toc12966855)

[4.1 数据预处理 8](#_Toc12966856)

[4.1.1 数据集的获取 8](#_Toc12966857)

[4.1.2 图像类别标注 8](#_Toc12966858)

[4.1.3 图像格式的转换 9](#_Toc12966859)

[4.1.4 数据集的划分 9](#_Toc12966860)

[4.2 图像分类与检测 10](#_Toc12966861)

[4.2.1 模型搭建 10](#_Toc12966862)

[4.2.2 模型训练 11](#_Toc12966863)

[4.2.3 模型优化 12](#_Toc12966864)

[4.2.4 模型评估 12](#_Toc12966865)

[4.3 可视化展示 12](#_Toc12966866)

[4.4 总结 14](#_Toc12966867)

一. 引言

1.1 编写目的

本项目的主要目的是实现图像的分类，拟采用深度学习的算法对图片进行分类与检测，为了提高图像分类与检测的准确率，在mask Faster R-CNN和SSD等众多图像分类算法模型中，尝试多种优化算法，激活函数，通过对比各种算法模型的优缺点，最终得出识别准确率较高的模型。

经过建模、图像数据预处理、深度学习实现图像检测分类、算法优化等多个步骤，最终实现对测试图片结构的可视化展示。

本文的目的在于为用户提供简单的说明。

1.2 项目背景

图像分类是作为图像视觉信息的主要表达方式，人们通过获取图片，经过筛选分类，最终可以实现快速的信息分类，提高了生产生活效率。目前，图像物体分类与检测已经成为计算机视觉领域的一个主要研究方向，它主要是一种利用计算机对图像进行处理、分析和理解，从而实现不同模式图像物体的分类。

随着大数据的发展，深度学习已经广泛应用于各个领域，2012年Hinton引入CNN解决ImageNet问题并取得巨大成功以来，ImageNet成为深度学习理论创新和技术突破的引擎，在计算机视觉领域掀起了一股研究热潮。这种基于深度学习的模型，其基本思想是通过有监督或者无监督的方式学习层次化的特征表达，来对物体进行从底层到高层的描述。

在大数据驱动的深度学习背景下，通过低层信号到高层特征的函数映射，来建立学习数据内部隐含关系的逻辑层次模型，以模仿人脑的视觉认知推理过程，从而使学习的特征具有更强的泛化能力和表达能力。因此基于深度模型架构通用性和统一性的视觉识别框架是当前计算机视觉领域的发展趋势，也是目前深度学习理论在视觉领域取得更深层次技术突破的一个切入点。

1.3 参考资料

[1] 黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6):1225-1240.

[2] 刘栋, 李素, 曹志冬. 深度学习及其在图像物体分类与检测中的应用综述[J]. 计算机科学, 2016, 43(12):13-23.

[3] 周俊宇, 赵艳明. 卷积神经网络在图像分类和目标检测应用综述[J]. 计算机工程与应用, 2017, 53(13):34-41.

[4] 陈荣, 曹永锋, 孙洪. 基于主动学习和半监督学习的多类图像分类[J]. 自动化学报, 2011, 37(8):954-962

[5] 张俊格. 基于视觉结构表达与建模的物体检测研究[D]. 中国科学院大学, 2013.

[6] Wang X, Thome N, Cord M. Gaze latent support vector machine for image classification[C]// IEEE International Conference on Image Processing. IEEE, 2016:236-240.

[7] Nielsen F, Muzellec B, Nock R. Classification with mixtures of curved mahalanobis metrics[C]// IEEE International Conference on Image Processing. IEEE, 2016:241-245.

[8] Chung A G, Shafiee M J, Wong A. Random Feature Maps via a Layered Random Projection (LaRP) Framework for Object Classification[J]. 2016.

[9] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]// Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2003:I-511-I-518 vol.1.

二. 软件概述

2.1 目标

图像分类系统旨在提出一种针对物体图像分类与检测的高效率算法，并对该算法进行可视化展示。系统主要使用谷歌的开源框架TensorFlow中的多个物体检测模型来进行图像的分类与检测，实现对每一张测试图片能够识别并检测出图片中物体的位置和类别。其中SSD模型将目标检测中的四个步骤（候选区域生成，特征提取，分类，位置精修）统一到一个深度网络框架内，该模型用Region Proposal Network（RPN）方法代替之前模型中生成检测框的EdgeBoxes方法或Selective Search方法，极大的提高了物体分类及检测的准确率和速度。

本系统通过使用SSD模型对自己的图像数据集进行训练和测试，最后将所检测图像的结果可视化展示出来。

2.2 功能

本项目的主要研究内容：

(1) 图像数据集的预处理：主要是在建模前对训练数据进行处理，本项目图像数据集的预处理主要包括对图像进行采集、图像格式进行转换、对采集的而图像进行标注、实心图像集的分类。

(2) 模型的构建及训练：本项目是基于TensorFlow框架，调用物体检测的多个模型，如SSD模型、R\_CNN模型、Faster R\_CNN模型，Mask R\_CNN模型等，将处理好的训练集数据放入模型中进行训练，训练过程中观察模型参数变化，待模型的损失函数收敛后，生成自己的模型文件。

(3) 模型的对比及优化：利用验证集对多个模型进行测试验证，通过调整模型中参数达到优化效果。

(4) 物体分类及检测：输入测试数据通过训练好的模型对输入的数据进行分类识别，最终时间图像的分类与检测。

(5) 物体的形态识别：在实现图像分类与检测的同时，可以对图像的体态进行识别，从而实现不同物体的体态识别。

(6) 图像检测结果的可视化展示

2.3 性能

1. 响应时间

图像分类系统尽可能实现端到端的实时查询。用户输入图片，即可在短暂的时间内输出检测结果。

2. 结果的准确性和合理性

图像分类系统对用户输入的图像中物体的位置和分类识别的准确率与已有的模型系统准确率相当。

3. 灵活性

系统接口清晰易读、符合接口规范，便于以后的扩充。

三. 运行环境

3.1 硬件

OSX 内存16G；

Windows7 内存8G。

3.2 支持软件

语言：Python

开发工具：Anaconda、Jupyter Notebook

技术平台：Tensorflow

技术API：Object-Detection API、Machine Learning API

四. 使用说明

4.1 数据预处理

数据预处理模块包括数据集的获取、图像格式的转换、图像类别标注和数据集的划分。图像分类检测模块包括模型的搭建、训练和评估三个部分。

4.1.1 数据集的获取

图像分类模型需要大量的数据来进行学习，我们需要从网络资源上下载多类别的图像数据集；利用脚本文件对下载好的图像统一命名，便于后面的使用方便。

其中，TensorFlow中SSD、Faster R-CNN、Mask R-CNN默认使用的数据集是VOC2012数据集。这个数据集以scene understanding为目标，主要从复杂的日常场景中截取，图像中的目标通过精确的segmentation进行位置的标定。图像包括91类目标，328,000影像和2,500,000个label。该数据集主要解决3个问题：目标检测，目标之间的上下文关系，目标的2维上的精确定位。

模型训练阶段使用的数据集是ImageNet数据集，Imagenet数据集有1400多万幅图片，涵盖2万多个类别。

4.1.2 图像类别标注

对下载的图像数据进行标签的标注，方便进行模型的训练。使用LabelImage工具，该工具使用python实现，使用简单方便。

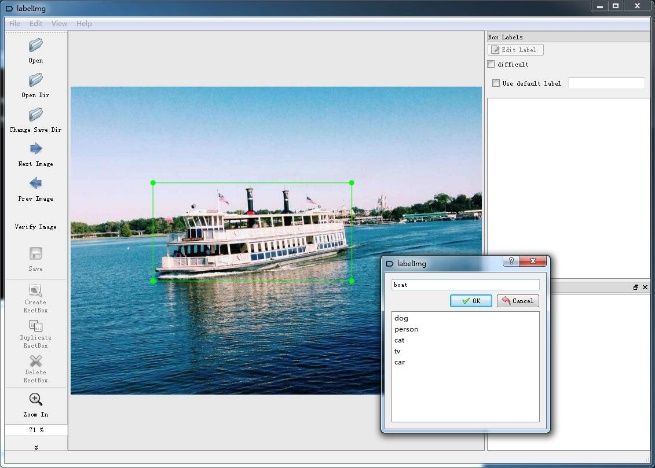


图4-1 使用软件标注图

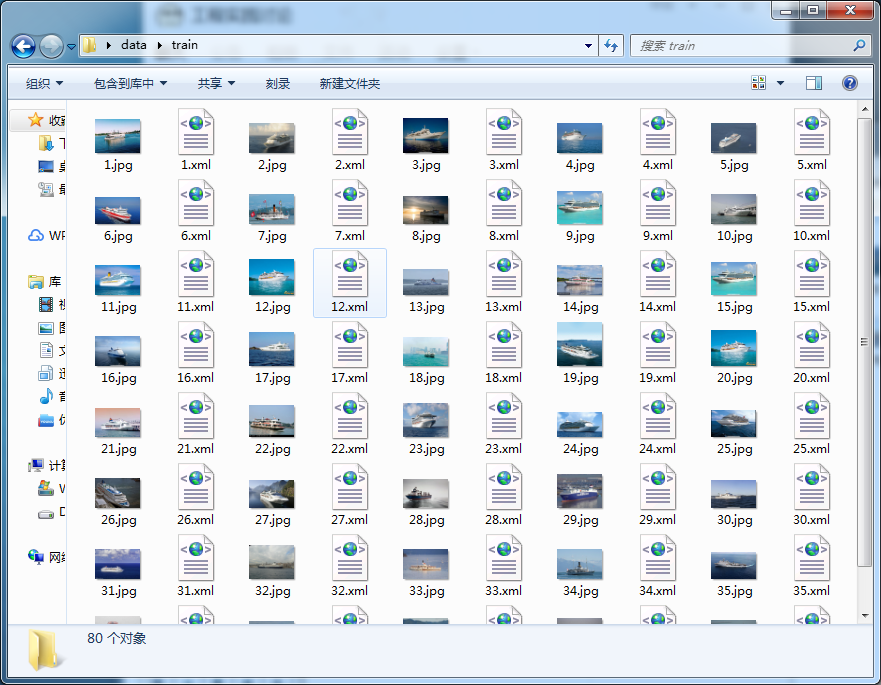


图4-2 标注后生成XML文件

LabelImg对图片标注后生成XML文件，该XML文件中包含了图片中物体的个数，类别以及物体的具体位置。

4.1.3 图像格式的转换

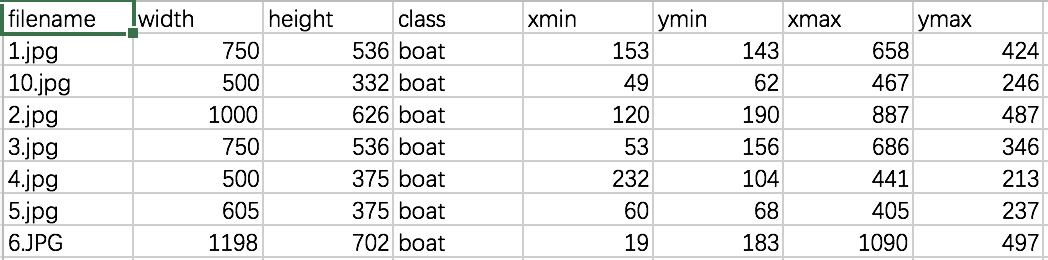
数据集需要符合API所需的TFRecord格式，而我们下载的数据集一般格式为PASCAL VOC格式，需要将其转换，之后得到如下图所示的CSV格式的数据集，作为我们最终的数据集。下图中，width和height分别表示图片的宽度和长度，class表示物体的类别；xmin、xmax、ymin、ymax表示物体的坐标。

图4-3 数据集格式图

4.1.4 数据集的划分

将数据集划分成训练集，验证集和测试集三个部分。使用训练集数据所有候选模型进行参数估计；使用验证集为检验样本，然后计算预测均方误差，比较各个模型的预测均方误差，选择预测均方误差最小的拟合模型为选择模型；测试集用于对模型的泛化能力的评估。

4.2 图像分类与检测

这个模块是分类系统的核心模块。主要通过调用谷歌开源TensorFlow框架中物体检测模型，通过训练集中的训练数据让模型进行学习；通过使用TensorBoard工具观察再利用使用验证集为检验样本，然后计算预测均方误差，比较各个模型的预测均方误差，选择预测均方误差最小的拟合模型为选择模型；最后利用测试集对模型进行测试并检验模型的泛化能力。

4.2.1 模型搭建

配置好TensorFlow环境以及物体检测所需的安装包，下载Faster-RCNN模型和SSD模型。并且模型进行配置，设定被识别物体的种类的pbtxt文件， 识别20种物体，设置训练步数。

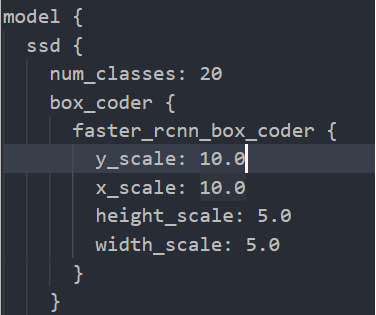


图4-4 配置参数

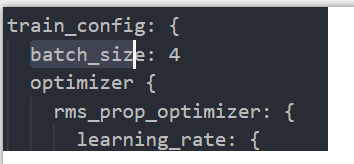


图4-5 设置训练步长

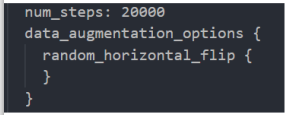


图4-5 设置训练步数

4.2.2 模型训练

将处理好的训练集数据放入模型中进行训练，训练过程中观察模型参数变化，待模型的损失函数收敛后，生成自己的模型文件，配置不同的训练步长以及训练次数。

主要使用TensorBoard来观察模型训练过程的参数变化。

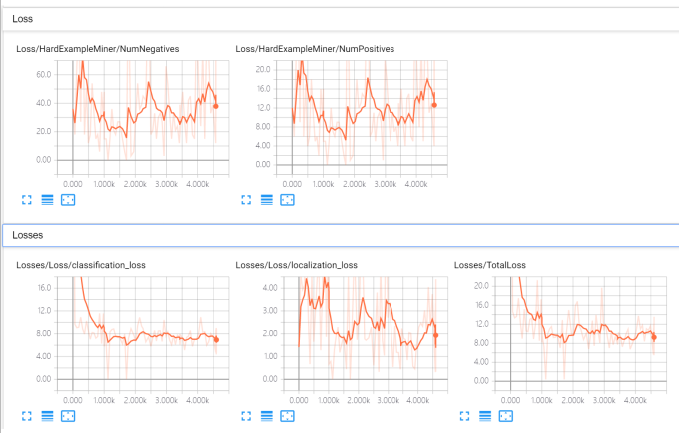
 

图4-6 batchsize为1训练变化（SSD） 图4-7 batchsize为4训练变化（SSD）

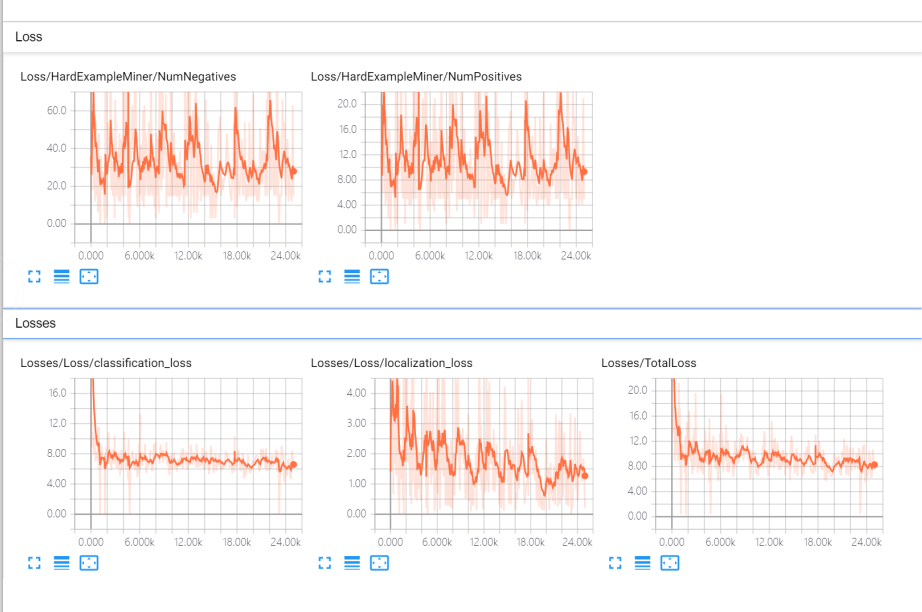
 

图4-8batchsize为24训练变化（SSD） 图4-9 batchsize为24训练变化（faster-rcnn）

主要观察TotalLoss这个损失函数的变化情况，如下图所示，训练一段时间后Faster-RCNN的损失函数明显下降。

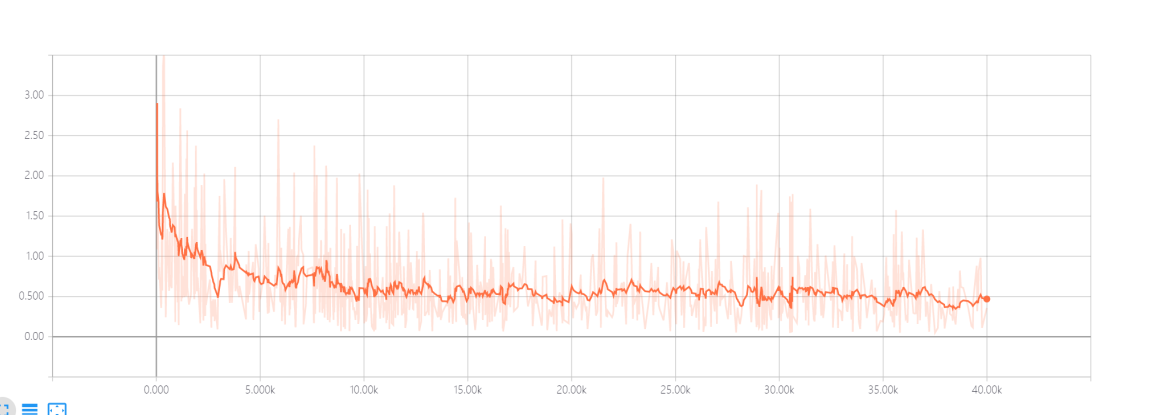


图4-10 训练过程中的TotalLoss的变化情况

观察模型参数收敛时，终止训练。生成一个pb文件，对pb文件进行处理后生成模型，也即后面使用的测试模型。

4.2.3 模型优化

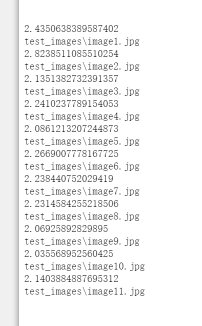
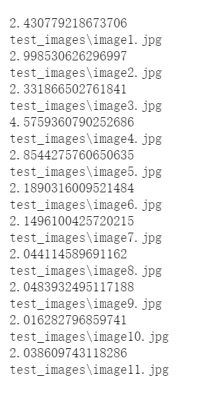
通过多次调整参数，生成多个自己训练的模型。利用验证集对多个模型进行测试验证，通过调整模型中参数达到优化效果。

4.2.4 模型评估

通过对的测试照片中物体类别检测正确与否以及可能性大小、物体检测框的准确度、检测的速度以及检测结果的准确性。

4.3 可视化展示

将图片放入指定文件夹中，在Jupyter Notebook中执行代码。可以查看模型检测速度和图片的检测结果。利用训练好的模型对用户传入的图片进行检测。对图片中的物体位置进行标注，以及物体类别预测的得分。

SSD（batchsize=4）速度 SSD（batchsize=24）速度 Faster-RCNN（batchsize=1）速度

图4-11 不同模型的训练速度

1. 使用batch\_size=4 step=25000的参数对模型训练，时长20小时（SSD)模型，检测速度2.43506s

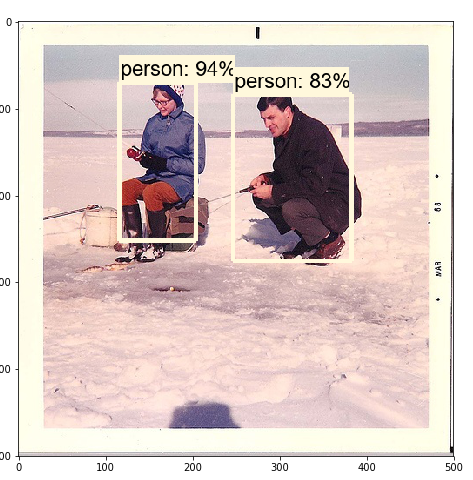


图4-12 模型检测结果

1. 使用batch\_size=24 step=25000的参数对模型训练，时长50小时（SSD)模型，检测速度2.99853s



图4-13 模型检测结果

1. 使用batch\_size=1 step=40000的参数对模型训练，时长50小时（Faster-RCNN)模型，检测速度6.811248s

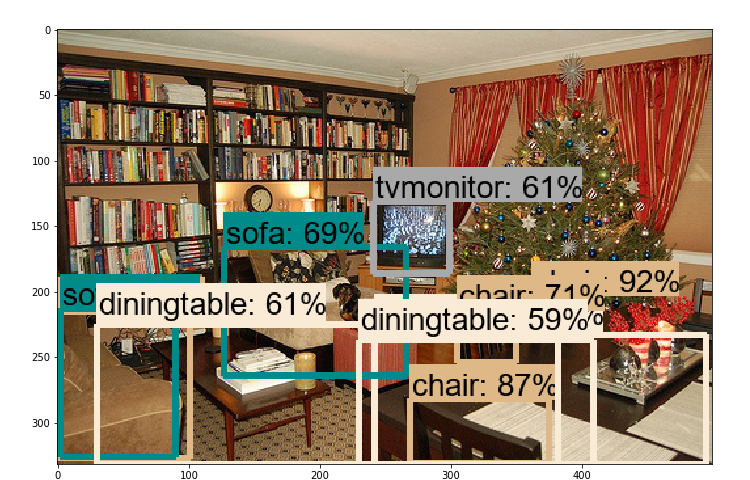
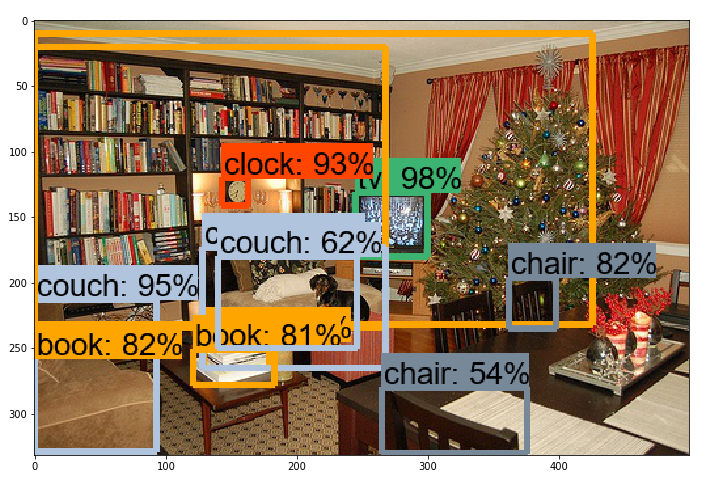
 

图4-14 模型检测结果图 4-15官方模型检测结果

4.4 总结

根据模型的检测速度和检测结果，可以得到以下结论：

1. Faster-RCNN检测速度较慢， 但是准确率相对SSD要高。 在训练过程中， loss稳定于0.5左右， 但是SSD\_mobilenet在训练过程中，loss一直维持在较高的水平。模型检测速度 57.974512s
2. 与官方模型相比， 因为设置的检测目标不同， 排除未设置的检测对象， 可以看到对于Faster-RCNN来说， 相同的物体检测准确率较高， 同时也可以检测到官方模型所未能检测出的对象。而SSD\_mobilenet， 官方模型准确率高于个人训练结果。
3. 从检测速度上来看， 官方SSD较慢， 但是准确率较高。官方Faster-RCNN与个人训练模型速度相比相近或稍慢。

因此个人训练的Faster-RCNN模型检测准确率更高。