**中国科学技术大学软件学院**

**软件工程实验项目环节**

**结题报告**

**项 目 名 称： 图像分类系统**

**成 员 名 单： 陈晨 范广宝 马鲜艳 王迪**

**导 师： 张曙**

**工 程 领 域： 图像分类**

**研 究 方 向： 物体分类与检测**

**开 题 时 间： 2018年11月28日**

**中国科学技术大学软件学院**

**填表日期： 2019 年 06 月30日**

小组成员贡献比：

|  |  |
| --- | --- |
| *小组成员* | *贡献率* |
| 陈晨 | 20% |
| 范广宝 | 25% |
| 马鲜艳  王迪 | 25%  30% |

# 一、 **简况**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **名称** | **中文** | **图像分类系统** | | | |
| **英文** | **Image Classification System** | | | |
| **项目组成员名单** | **姓名** | | **学号** | 项目中的分工 | 签 章 |
| **陈晨** | | **SA18225027** | **tfrecord格式转换， 测试图片标记** |  |
| **范广宝** | | **SA18225070** | **ssd mobilenet模型搭建** |  |
| **马鲜艳** | | **SA18225277** | **模型优化、测试** |  |
| **王迪** | | **SA18225352** | **faster rcnn模型搭建** |  |
| **中 英 文 摘 要** | 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。  本项目利用已有的对物体图像分类与检测的算法，进行分析后加以改进，提高算法效率。使用开源框架TensorFlow进行图像的检测与分类，对于输入的图像可以进行分析，对图像中物体进行检测、分类及标识。利用SSD模型统一目标检测过程，为了提高图像物体分类与检测的时间效率以及算法的准确度，在SSD与R-CNN等多个算法中，尝试多种优化算法，激活函数，并且对比多种算法模型的优缺点，得出准确率较高的模型，最终实现图形检测分类结果的可视化展示。  Image classification is the main expression of visual information. People can obtain fast image classification and improve production and living efficiency by obtaining images and filtering and sorting. At present, image object classification and detection has become a major research direction in the field of computer vision. It is mainly a kind of computer processing, analysis and understanding of images, so as to realize the classification of image objects in different modes.  This project utilizes the existing algorithms for classification and detection of object images, analyzes them and improves them to improve the efficiency of the algorithm. The open source framework TensorFlow is used to detect and classify images, and the input images can be analyzed to detect, classify and identify objects in the images. Using the SSD model to unify the target detection process, in order to improve the time efficiency of image object classification and detection and the accuracy of the algorithm, in multiple algorithms such as SSD and R-CNN, try multiple optimization algorithms, activate functions, and compare various algorithms. The advantages and disadvantages of the model are obtained, and the model with higher accuracy is obtained, and finally the visual display of the classification result of the graphic detection is realized. | | | | |
| **主题词** | **主题词数量不多于三个，主题词之间空一格（英文用“/ ”分隔）** | | | | |
| **中文** | **目标检测/深度学习/神经网络** | | | |
| **英文** | **Object detection/deep learning/CNN** | | | |

# 二、选题依据

|  |  |  |
| --- | --- | --- |
| 1. **阐述该选题的研究意义，或工程设计的价值和意义，国内外概况和发展趋势，选题的先进性和实用性，技术难度及工作量。**   **选题的研究意义：**  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。  **国内外概况和发展趋势：**  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。  首先以国际视觉算法的相关竞赛为主线对图像物体分类与检测算法近年来的主要进展进行综述。该算法最初产生于自然语言处理领域，通过建模文档中单词出现的频率来对文档进行描述与表达。以底层特征提取、特征编码、特征汇聚、空间约束、分类器设计、模型融合作为标准物体分类框架。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。  现如今，物体检测问题更侧重于物体结构信息，物体检测的输入是包含物体的窗口，所以学术界根据窗口位置获取方法的不同，将物体检测方法划分为滑动窗口和广义霍夫投票两类。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。  **选题的先进性和实用性：**  图像识别已经成为当下社会中的一个热点话题，对着大数据时代的到来，图像识别结合机器学习已经广泛应用与各个行业，如：监控、人脸识别签到、图片搜索等多个方面，同时也为人们的生活带来了很多便利。图像识别通过图像的检测技术实现图像数据的获取，结合深度学习，是机器更加智慧的对图像进行分类，这不仅减轻了人工实现图像分类的任务，而且提高了图像分类的准  确率和速度，具有很大的应用前景。 | | |
| **技术难度及工作量**  （1）模型复杂度高，优化困难。神经网络的容量没有上限，表达能力非常强，这是它的一个重要的优点。另一方面也对模型的优化造成了非常大的困难，网络越复杂，模型的能量面越高低不平，到处是极小点。研究模型初始化方式、优化算法，提高神经网络的判别能力，是深度学习的一个重要研究内容。  （2）解释性差。层次化表达在视觉皮层理论和函数论等方面具有其理论依据，然而，在实际应用中，学习到的模型通常没有很好的解释性。  （3）计算强度高。目前虽然每层是高度并行化的前馈网络，但是计算强度还是比较高，可能需要采用GPU等硬件来完成。  （4）模型缺少结构约束。深度学习模型通常只对网络的“输入输出”进行建模，却缺少必要的结构先验的约束。  **2. 参考文献。**   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. 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 | | |
| 3.1课题内容  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，最后将我们所检测图像的结果可视化展示出来。 3.2系统需求分析 **3.2.1 系统角色**  图像分类系统主要面向个人用户，管理员用户作为系统后台人员参与系统维护。图3.2为系统角色图。    图3.2 系统角色  **3.2.2 系统用例分析**  **1. 个人用户用例**  **个人用户**可以使用以下功能：  (1)输入图片；  (2)查看图片识别结果。  **2. 管理员用户用例**  **管理员用户**可以使用以下功能：  (1)输入图片数据集；  (2)查看图片识别结果；  (3)对比不同模型识别结果；  (4)调用模型；  (5)训练模型；  (6)评估模型。    图3.4 管理用例图 3.3 系统概要设计 **3.3.1 系统结构设计**  图像分类系统主要功能是实现对测试图片能够识别并检测出图片中物体的位置和类别。系统总体结构图如下图所示。    图3.5图像分类系统总体结构图  **3.3.2 功能模块设计**  该系统的关键模块由三大部分组成：数据预处理模块，主要包括对图像数据集的获取、数据集格式的转换以及图像物体的类别标注和划分；图像分类检测模块，主要包括模型的搭建训练以及测试评估；以及可视化展示模块。    图3.6 图像分类系统模块图  **1.数据预处理模块**  数据预处理模块主要实现输入数据的标准化。因为系统所处理图像有统一的格式要求，输入的图像应经过预处理，达到可以被系统处理的要求。主要功能有：数据集获取、图像格式转换、数据集划分及图像类别标注。  **（1）数据集获取**  从网络上获取合适的数据集，对数据集进行统一命名以便后续工作。  **（2）格式转换**  TensorFlow使用官方推荐的数据读取标准格式：tfrecord。tfrecord是一种同时存放图像数据和标签的二进制数据格式，有利于使用内存及数据的使用。将获得的数据集保存为tfrecord格式。  **（3）图像类别标注**  对图像进行分类标注。使用labelimage标注工具，用矩形框进行标注，标注出图像中物体的位置及名称。  **（4）数据集的划分**  模块调用TensorFlow中目标检测模型，使用训练集训练模型，使用验证集检验样本，计算均方误差进行各模型的比较，从中选出误差最小的模型为结果。最后用测试集数据对模型进行评估。  **2.图像分类及检测模块**  图像分类检测模块调用TensorFlow中目标检测模型，使用训练集训练模型，使用验证集检验样本，计算均方误差进行各模型的比较，从中选出误差最小的模型为结果。最后用测试集数据对模型进行评估。  **（1）模型搭建**  配置好TensorFlow环境，利用pip安装好所需的包，为调用目标检测模型、使用训练集对被选中的模型进行训练做好准备。  **（2）模型训练**  将已被处理的数据集输入模型进行训练，训练结束后得到训练后的模型。  **（3）模型的优化**  利用验证集对模型进行检验，调整参数进行模型优化。  **（4）模型的评估**  使用测试集对模型进行测试，主要测试检测准确性、速度等。  **3.可视化展示模块**  可视化界面展示模块对项目结果进行可视化展示。 3.4.1**模型训练接口设计** 模型搭建完成后，针对训练集的数据进行训练与测试，以下是训练接口的详细设计。    图3-7 接口模块图  **1）learn\_runner**      **2）Estimator（估算器）类**  Estimator类代表了一个模型，以及如何对这个模型进行训练和评估。    **3）Experiment（实验）类**  Experiment类定义了如何训练模型。    **4）Train**  Train是模型的训练方法。    **5）Evaluate**  Evaluate是在模型训练时评估模型训练的方法。   3.5完成情况 **（1）数据预处理模块完成情况**  在图像分类系统中，数据预处理模块主要是从现有的网络资源中搜索图像数据集，并对图像进行处理使之转换成能被后续模型识别的照片格式。该模块主要包含图像数据集的下载、统一命名、格式转换及图像类别标注模块。  **1）数据集获取**  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。主要使用imagenet的voc2012数据集。    图3-8 数据集结构  **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。修改后的工具支持多label的标签标注。    图3-9 软件标注成果图  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。  **3）格式转换**  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3-10 数据集格式图  **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。   1. **图像分类检测模块完成情况**   **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SSD模型。 对SSD mobilenet和faster\_rcnn模型进行不同的配置。    图3-11 配置参数    图3-12 配置参数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3-13 训练过程中的参数变化  主要观察TotalLoss这个损失函数的变化情况，如下图所示，训练一段时间后损失函数明显下降。    图3-14 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。  **3）模型的优化**  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）模型的评估**  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。  1. 不同模型间结果对比  ssd mobilenet模型检测速度快于faster\_rcnn，性能足以用于实时性要求较低的实时视频检测。而faster\_rcnn检测速度慢于ssd mobilenet，但在准确性上强于ssd mobilenet。如果要求ssd\_mobilenet在准确性上有所提高，需要对具体的参数进行调整，且将消耗更多的时间在模型的训练上。    图3-15 不同模型运行速度对比（左：ssd，右：faster\_rcnn  相较于Faster R-CNN和Mask R-CNN，SSD在单一种类有重叠物体的图像中识别能力较弱，如下列图所示。Mask R-CNN在复杂图像中识别种类更多，还增加了物体轮廓标注，如图3-16所示。  mask1  图3-16 Mask R-CNN  ssd  图3-17 SSD  faster2  图3-18 Faster R-CNN  2、同一模型不同参数结果对比  分别将batch\_size设置为1，4，24，并在不同情况下设置不同的num\_steps，进行性能的对比。faster\_rcnn无法修改batch\_size，所以对比基于ssd mobilenet模型。  根据测试结果，num\_steps相同时，batch\_size越高，得到的结果越准确；而在batch\_size相同时，在一定程度上，num\_steps越大，准确度越高，但在参数调整不合适的情况下，训练过程中可能会陷于局部优化，loss不会持续下降至可以接受的范围，这也意味着准确度的不足。    图3-19 batch\_size为24时 图3-20 batch\_size为4时  3、同官方结果对比  使用ssd\_mobilene模型时，由于陷于局部优化，结果不尽人意，与官方模型对比准确度低下。而个人训练的faster\_rcnn模型对测试图片进行测试的结果如下，与官方训练的模型进行对比。考虑到官方训练的模型使用的标签与个人训练时使用的不同，所以对于同一物体测试出的结果不同仅仅是“名字”的不同。忽略这一点，可以看出官方模型在检测标签内都包含的对象时，个人训练的模型精度高于官方模型，且在测试同一图片时，个人训练的模型检测速度更高。    图3-21 项目模型检测结果    图3-22官方模型检测结果  **（6）采用的开发方法、环境，测试方案等**  **开发平台：**OSX 内存16G；Windows10 内存24G。  **语言：**Python  **开发工具：**Anaconda、Jupyter Notebook  **技术平台：**TensorFlow  **技术API：**Object-Detection API   3.6 总结与展望 3.6.1总结  物体分类与检测是计算机视觉、模式识别与机器学习领域非常活跃的研究方向。物体分类与检测在很多领域得到广泛应用，包括安防领域的人脸识别、行人检测、智能视频分析、行人跟踪等，交通领域的交通场景物体识别、车辆计数、逆行检测、车牌检测与识别，以及互联网领域的基于内容的图像检索、相册自动归类等。  图像作为视觉信息的主要表达手段，是人类感知客观世界的主要信息来源，如何快速且有效地定位出图像中的主要信息内容，已经成为了人们研究的热点。在传统视觉领域，物体检测是一个非常热门的研究方向。受有限应用场景的影响，物体检测直到上个世纪90年代才开始逐渐走入正轨。物体检测对于人眼来说并不困难，但对于计算机来说，面对的是RGB像素矩阵，很难从图像中直接得到这样的抽象概念并定位其位置，再加上物体姿态、光照和复杂背景混杂在一起，使得物体检测更加困难。  物体分类与检测的研究，是整个计算机视觉研究的基石，是解决跟踪、分割、场景理解等其他复杂视觉问题的基础。欲对实际复杂场景进行自动分析与理解，首先就需要确定图像中存在什么物体（分类问题），或者是确定图像中什么位置存在什么物体(检测问题)。鉴于物体分类与检测在计算机视觉领域的重要地位，研究鲁棒、准确的物体分类与检测算法，无疑有着重要的理论意义和实际意义。  通过这段时间的设计与开发，总体上完成了以下几项工作：  1.研究了深度学习中图像分类与检测技术，如TensorFlow、SSD、Faster R-CNN、Mask R-CNN等模型，参考了一些相关的技术论文和文献资料，对深度学习模型的训练和评估有了一定的认识和了解。  2.研究了开源框架TensorFlow等技术，熟悉了Anaconda、Jupter Notebook等工具的使用，从而为系统的实现奠定好基础。  3.在研究设计的基础之上对系统进行实现，该系统大体上实现了总体设计方案中所提到的数据预处理模块、图像分类与检测模块和可视化展示模块等3个模块的功能。  4.对系统进行测试，尽管在测试过程中还有一些尚未解决的问题，但是这些都是预期可以解决的。  3.6.2展望  图像分类系统的设计与实现涉及多方面的理论、方法和技术。本系统还有许多新的问题需要解决，需要在实际应用中不断积累和完善，在以下几个方面，还需要做进一步的研究和开发。  1.论文只考虑了系统的几项基本组成，并未在扩展功能等方面进行深入发掘。使得系统只考虑了一些基本功能，如何加入扩展功能，增加用户粘性，是我们后续需要关注的问题。  2.对于ssd mobilenet模型，由于参数调整有误，在多种情况下，模型在训练过程中都先后陷入局部优化，导致最后导出的模型准确度较低，后续过程中可以深入了解模型，找出进一步降低loss的方法。 |
|  |

四、工作进度完成情况

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4.1 工作进度 开题阶段：（11月底之前，**已完成**）  本阶段抓紧完成项目的前期准备工作，本阶段结束后项目组将提交《软件工程实验项目开题报告》并进行开题答辩。   |  |  | | --- | --- | | 主要任务：  完成前期技术预研  选定项目组织结构  完成项目的结构分解  完成项目工作量估算  完成项目进度估算  完成项目需求分析  完成确认测试技术  确认过程模型 | 阶段工作产品：  开题答辩PPT  需求定义文档  SRS规格说明文档  确认测试计划文档  技术预研报告  项目管理计划文档 |   设计开发阶段：（2017年11月～2018年3月底，**已完成**）  本阶段将完成整个项目的总体设计和详细设计，并进行编码   |  |  | | --- | --- | | 主要任务：  完成体系结构设计  完成详细设计  完成单元测试计划  完成编码规范  完成核心算法的设计  完成核心组件编码开发 | 阶段工作产品：  软件工程师兼中期检查报告  中期检查答辩PPT  设计文档  测试计划文档  编码规范说明文档  核心组件代码 |   后期开发阶段：（2018年4月～2018年6月，**已完成**）  本阶段将完成所有编码工作，完善过程中所产生的文档，并对项目进行最后测试   |  |  | | --- | --- | | 主要任务：  完成项目的所有编码工作  功能测试  项目的配置和部署  撰写用户使用手册  撰写项目技术论文  结题答辩 | 阶段工作产品：  软件工程试验结题报告  结题答辩PPT  测试报告  项目安装或配置手册  项目源代码  项目技术论文  小组成员贡献比说明 | |  |  | |
| **导师意见（对选题和工作过程及成果进行说明，并给出成绩。）**  **导师签名：**  **年 月 日** |
| **答辩小组意见**  **答辩组长签名：**  **年 月 日** |