

车辆大规模精准搜索项目报告

杨国峰

硕士

张越

硕士

一、摘要

车辆大规模精准搜索（以下简称车辆检索）在实际应用中具有非常重要的意义。与其他对象检索任务类似，车辆检索任务可以定义为：给定两部分图片数据 ref(车辆图片数据库)和 query(测试车辆图片)，目标是对 query 中每张测试图片在 ref 中找出所有属于相同车辆的图片。在课程项目中，我们的任务是对 query 中每张测试图片在 ref 中找出所有属于相同车辆的图片并排在查找结果的前面，采用 MAP@K 方法对算法结果进行评测。

我们将车辆大规模精准搜索的任务划分为车辆型号(model)识别任务和车身颜色识别任务这两个子任务。在车辆型号识别任务中，我们使用了基于迁移学习的深度卷积神经网络模型算法来判断车辆型号。使用的深度卷积神经网络模型包括 VGG16、Inception_V3 和 ResNet50 等，将这些模型进行迁移学习，能使其在训练集上达到 97%的型号正确分类效果；车身颜色识别任务使用了基于 OpenCV 的 HSV 颜色识别算法。将两个子任务结合就可以实现车辆精准识别的任务，我们的算法在测试集上达到非常优异的识别效果。

二、项目意义

图像中特定目标的识别一直是计算机视觉领域的研究热点,车辆图像识别不论在民用还是军用方面都具有重要意义。在民用方面,车辆识别推进了智能交通、电子收费、车辆监管、安防等领域的应用;在军用方面,车辆识别对战场环境中车辆(坦克、步兵战车等)的部署动态,识别与跟踪,精确打击等方面都发挥着关键作用。

城市化进程的推进、汽车保有量的增加、城市交通及车辆出入口管理要求的提高,互联网、物联网、云计算、大数据技术的发展,推动着人工智能技术在车辆识别场景中的快速发展。

尽管有如此迫切的现实需求,车辆精准搜索任务在实际应用中依然存在不少挑战。它们主要来自于四个方面:

1、车辆所处的背景环境复杂多变,如何在识别过程中过滤图片背景信息,使算法只关注于车辆本身是一大挑战;

2、车辆图像易受光照、天气等外界因素干扰,造成车身模糊或图片整体模糊的现象,如何消除这些干扰是另一大挑战;

3、车辆图像拍摄角度和高度不同会导致车辆出现一定程度的形变和尺度变化,如何克服这些变化也是一个挑战;

4、车辆品牌、型号、颜色众多,不同品牌车辆存在外形相似的情况、相同型号车辆存在颜色不同的情况。精准的车辆识别要求我们建立庞大的车辆数据库,并能对这些车辆特征进行有效的提取;

为了解决上述问题,人们提出了许多算法。这些算法中有些在解决某一特定问题场景有非常优秀的表现。我们设计的识别算法,就是在吸收不同算法的优点的基础上实现的。

三、研究现状

要解决复杂环境中车辆识别的问题，首先需要了解车辆识别流程。车辆图像识别通常分为两个阶段：

1、样本训练阶段：对大量标签样本进行图像预处理、提取目标特征后，建立车辆目标特征集，再采取相应的学习算法进行分类训练。

2、图像目标识别阶段：按照图像预处理、提取目标特征、目标分类识别的步骤、最终确认测试图像中车辆属性。

（一）图像预处理

图像预处理是为了增强图像的可续性。外部环境因素和成像系统自身噪声的干扰，是 CCD 摄像机和图像采集卡获取静态图像的主要降质原因。针对系统自身噪声，常用的方法有滤波、灰度变换、二值化、图像复原、图像增强、形态学处理等。

1、背景环境处理

对背景环境处理是为了背景与车辆有效分离，主要包括车辆检测和背景去除两部分。

背景检测方法有：基于模型、基于背景属性。

模型方法：利用有关场景、目标、光照的先验知识构建背景模型；

属性方法：通过分析背景亮度、梯度、色彩、纹理等信息，对背景区进行标识。具普适性，逐渐成为主流方法。

背景去除的目的是将背景区的颜色、亮度和纹理与车辆区域状态进行区分，再将其去除。常用到的方法有：基于泊松方程的背景去除、基于梯度域的背景去除、使用子区域匹配亮度转移的背景去除等。

2、光照和天气处理

光照和天气会造成图像中车辆对比度降低，边缘轮廓信息模糊，给车辆识别带来严重干扰。

通常采用基于图像增强和基于物理模型两类方法。图像增强优势在于图像处理算法较为成熟，其原理是改进常用的图像增强算法来调整图像亮度或对比突出局部细节信息，能从某些角度有限的改善图像品质并获取有价值的信息；缺点是由于没有考虑图像降质的本质原因，图像或多或少会出现失真或信息丢失的情况。物理模型是基于大气退化模型或先验知识来复原退化图像，能更贴近的还原真实场景，对复杂场景的图像处理效果较好。

当前，一些优秀的算法都是在基于偏微分方程、深度关系、先验知识等算法上进行改进，取得了较为理想的效果。

3、角度和高度处理

拍到的画面中经常会出现车辆被物体或车辆间相互遮挡的情况，对车辆识别形成很大的干扰。解决地物遮挡的方法主要有特征模型、统计模型、推理和三维模型。特征模型和统计模型的方法应用最为广泛。

特征模型：算法运算较为复杂，易受环境中不相关物体的干扰；统计模型：仅依赖统计上的差异，可靠性不高；推理模型：对车辆轨迹、位置等先验信息依赖性强，相对简单的场景下处理遮挡效果优良，复杂场景下效果仍有待提升。

4、车辆品牌、型号、颜色处理

针对车辆大规模识别以前多以对车辆的车牌进行识别,同时通过比对交通运输部门的数据库提供的信息来进行车辆的识别。随着图像处理、拍摄和传输技术的发展,也只能进行小规模的车辆检索识别。

主要的方法:基于车牌的品牌识别,灰度阈值分割法车型识别,图像背景插值法车型识别,基于色差和彩色归一化的车身颜色识别,基于去雾和高光修复的车辆颜色识别。

(二) 车辆特征提取

车辆识别中常用的特征有:视觉特征、全局不变量特征和局部不变量特征。

视觉特征包括:灰度、亮度、对比度、几何特征(形状、长宽比)等;全局不变量特征包括:矩不变量和傅里叶描绘子等;局部不变量特征包括:角点、SIFT 特征和轮廓描述等。ps: 由于复杂环境背景下存在噪声较大、局部遮挡等情况,利用局部不变量特征进行识别是非常有效的方法。

特征提取典型的方法有:主分量分析、独立分量分析、Fisher 线性鉴别分析、因子分析和奇异值分解以及流形学习方法等。

按提取手段不同分为人工特征提取和自动学习特征提取。

1、人工特征提取

过去几十年,人工特征提取在图像模式识别领域一直占据主导地位,其主流思想是先采用区域选择方法定位提取感兴趣的区域,然后手工设计提取目标特征。

车辆特征提取建立在车辆检测的基础上,先搜索定位包含车辆的区域,再提取有用的车辆特征信息。

传统的车辆检测方法采用基于滑动窗口的穷尽策略对整幅图像遍历,通过设置不同的尺度、长宽比,搜索包含目标所有可能出现的位置。近年来,以 AdaBoost 框架和 DPM 框架的各种改进算法堪称佳作,取得了较好的效果。

手工车辆特征提取方法存在两个问题:(1) 基于滑动窗口的区域选择方法缺乏针对性,时间复杂度高,且窗口冗余;(2) 手工设计的特征对复杂环境背景中车辆特征的鲁棒性有待提高。

2、自动学习提取特征

从大量的训练样本图像中自动学习提取特征,神经网络一度被给予厚望。但其发展出现了瓶颈:训练容易出现拟合、算法复杂、训练速度慢。近年来,伴随着 GPU 和深度学习的快速发展,弥补了传统神经网络的不足。

深度学习的本质是通过多层非线性变换,从大规模数据集中自动学习特征。深层的结构使其具有极强的表达能力和学习能力,尤其擅长提取复杂的全局特征和上下文信息。

得益于卷积神经网络和候选区域算法,以 R-CNN 为代表的目标检测方法开启了深度学习应用于目标检测的先河。

以 RP+CNN 分类的目标检测框架主要基于 3 个步骤:步骤(1) 通过滑动窗口或 object proposals 方法产生大量的候选窗口;步骤(2) 使用训练好的分类器对候选窗口进行分类并对其进行打分;步骤(3) 采用非极大值抑制消除重叠检测。

鉴于卷积神经网络对通用物体分类识别非常有效,并在 ImageNet 图像分类和人脸识别、行人检测等方面取得了重大突破,如果将其应用于车辆识别,尤其在复杂环境背景下,将更具灵活性和普适性。

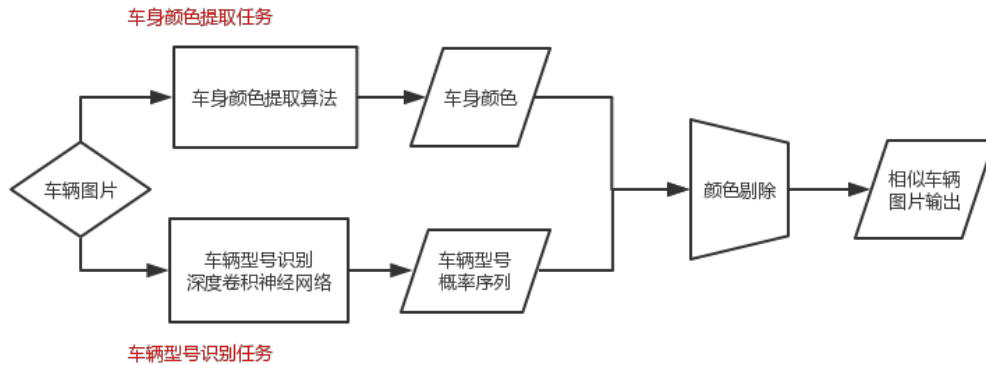
(三) 分类型器设计

分类器是解决目标识别的有效工具。常用的分类器有最小距离分类器、贝叶斯网络分类器、人工神经网络分类器、支持向量机等。目前应用最为广泛的分类器是神经网络分类器和支持向量机。

综上:实现车辆与复杂环境的分离,仍面临着很多挑战性的问题。随着数据的不断积累,技术的发展,场景应用将越发广泛普遍。

四、模型及算法详细说明

我们将车辆大规模精准识别的任务划分为两个子任务：车辆型号识别任务和车身颜色识别任务，将这两个子任务结合就可以实现车辆的识别。算法工作的具体流程如下图所示：



在算法开始之前，我们需要对数据进行预处理。根据项目说明，我们将 ref 数据库中的所有图片按照 modelID（即车辆型号）进行分类整理，使型号相同的车辆归入同一类中。预处理的操作有两个作用，一是为训练车辆型号识别网络做数据准备；二是方便算法的后续处理。

算法接收车辆图片信息作为输入，分三步实现车辆的精准识别。

第一步是使用车身颜色提取算法提取车身颜色。这个任务可以使用很多方法实现，例如机器学习方法或者基于规则的颜色提取算法等；

第二步是车辆型号识别。用深度卷积神经网络计算出输入车辆型号对训练集中所有车辆型号的概率分布，概率分布的大小关系代表了不同的车辆型号与输入车辆型号的相似程度关系，概率越大表示该型号与输入车辆型号越接近；

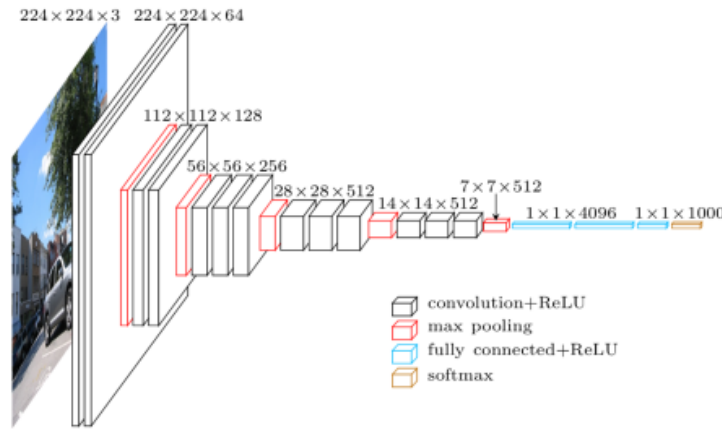
第三步是颜色剔除。算法在 ref 数据库中，按照型号相似度从高到低的顺序依此访问各个型号类别，在每个型号类别中，计算该类别中的所有车辆图片的颜色信息并剔除颜色不同的车辆图片，遍历访问所有类别直至达到某个结束条件（比如收集了足够多张车辆图片），然后输出结果。

（一）车辆型号识别任务

在车辆型号识别任务中我们使用了基于迁移学习的深度卷积神经网络模型的算法。卷积神经网络是一种前馈神经网络，它的神经元可以响应一部分覆盖范围内的周围单元，由于其局部敏感和方向选择的特性，其网络结构复杂度大大降低，在大型图像处理应用中有着出色的表现。在项目中我们具体用到的深度卷积神经网络模型有三种：VGG16、Inception_V3 和 ResNet50。这三种网络模型在图像分类问题中有着广泛的应用，尤其在一些特定的问题场景中有着非常优秀的性能表现，在下文中我们将以 VGG16 网络为例，介绍项目中使用的网络结构。

1、VGG16 卷积神经网络模型

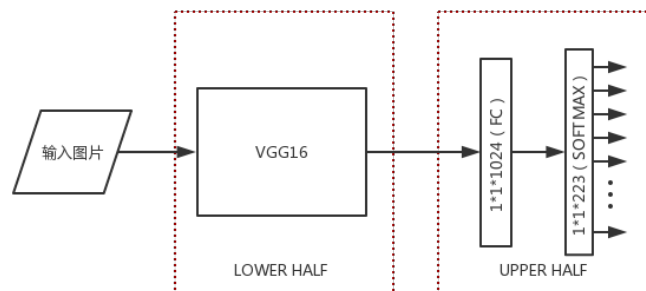
VGG 是由牛津大学的 K. Simonyan 和 A. Zisserman 等人提出的卷积神经网络模型，模型的设计之初就是为了实现大规模图像的识别任务。VGG16 网络模型的结构如下图所示：



网络的前半部交叉包含 13 层卷积层和 5 层池化层用来进行特征提取，后半部包含 3 个全连接层和 1 层 softmax 输出层进行结果的分类。VGG16 网络在 ImageNet 分类任务中达到了 92.7% 的 top-5 准确率，是一个非常基础、非常优秀的深度卷积神经网络模型。

2、基于迁移学习的深度卷积神经网络模型

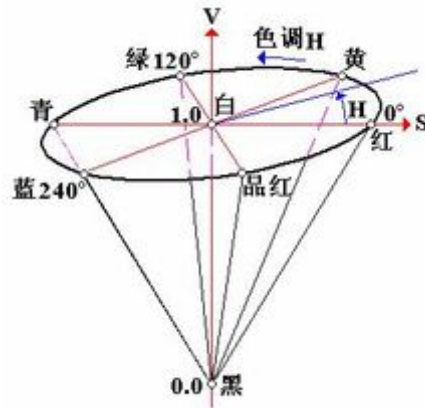
VGG16 网络非常深，参数数量非常多，因此如果直接将 VGG16 网络用于车辆型号识别任务，那么由于训练集样本数相对参数数量过少，会导致严重的过拟合问题。同时我们注意到，卷积神经网络的前半部分在图像识别任务中扮演着特征提取器的角色，不同的识别任务特征提取器的差别不大，这意味着特征提取器可以作为一个通用的结构应用在不同的图像识别任务中。于是我们采用了迁移学习的方法，将 VGG16 网络在 ImageNet 上训练的模型应用到我们的项目中来，得到的用于车辆精准搜索的网络结构如下图所示：



在网络的 UPPER HALF 中我们加入了 $1 \times 1 \times 1024$ 的全连接层，之后再连接一层 $1 \times 1 \times 223$ 的 softmax 层进行分类。

(二) 识别出图片中车辆的颜色(colorID)：使用 opencv 库函数进行处理

HSV(Hue, Saturation, Value)是根据颜色的直观特性由 A. R. Smith 在 1978 年创建的一种颜色空间，也称六角锥体模型(Hexcone Model)。这个模型中颜色的参数分别是：色调 (H)，饱和度 (S)，明度 (V)。



- 1、定义 HSV 颜色字典，将图片颜色转为 HSV
- 2、使用 `cv2.inRange()` 函数进行背景颜色过滤
- 3、将过滤后的颜色进行二值化处理
- 4、进行形态学腐蚀膨胀，`cv2.dilate()`
- 5、统计不同颜色区域面积，如果使用某种颜色分量过滤背景后，出现车辆的轮廓且可视面积最大，则可认为车辆为该颜色

(三) 计算出图片的直方图重合程度/汉明距离(Hamming Distance)并按序检索出结果

aHash.py (平均哈希算法): 基于比较灰度图每个像素与平均值来实现的, 最适用于缩略图, 放大图搜索。dHash.py (差异哈希算法): 相比 pHash, dHash 的速度要快的多, 相比 aHash, dHash 在效率几乎相同的情况下的效果要更好, 它是基于渐变实现的。pHash.py (感知哈希算法): 平均哈希算法过于严格, 不够精确, 更适合搜索缩略图, 为了获得更精确的结果可以选择感知哈希算法, 采用的是 DCT (离散余弦变换) 来降低频率的方法。

- 1、aHash.py——用平均哈希法实现识别相似图片
 - ①.缩放图片, 一般大小为 8×8 , 64 个像素值
 - ②.转化为灰度图
 - ③.计算平均值: 计算进行灰度处理后图片的所有像素点的平均值, 直接用 numpy 中的 `mean()` 计算即可
 - ④.比较像素灰度值: 遍历灰度图片每一个像素, 如果大于平均值记录为 1, 否则为 0
 - ⑤.得到信息指纹: 组合 64 个 bit 位, 顺序随意保持一致性
 - ⑥.比对两张图片的指纹, 获得汉明距离
- 2、pHash.py——通过对图片矩阵进行离散余弦变换来识别相似图片
 - ①.缩小图片: 32×32 是一个较好的大小, 这样方便 DCT 计算
 - ②.转化为灰度图
 - ③.计算 DCT: 利用 Opencv 中提供的 `dct()` 方法, 注意输入的图像必须是 32 位浮点型, 所以先利用 numpy 中的 `float32` 进行转换
 - ④.缩小 DCT: DCT 计算后的矩阵是 32×32 , 保留左上角的 8×8 , 这些代表的图片的最低频率
 - ⑤.计算平均值: 计算缩小 DCT 后的所有像素点的平均值
 - ⑥.进一步减小 DCT: 大于平均值记录为 1, 反之记录为 0

- ⑦.得到信息指纹：组合 64 个信息位，顺序随意保持一致性
- ⑧.比对两张图片的指纹，获得汉明距离

3、dHash.py——利用差异值哈希算法进行识别

- ①.缩小图片：收缩到 9*8 的大小，以便它有 72 的像素点
- ②.转化为灰度图
- ③.计算差异值：dHash 算法工作在相邻像素之间，这样每行 9 个像素之间产生了 8 个不同的差异，一共 8 行，则产生了 64 个差异值
- ④.得到信息指纹：如果左边的像素比右边的更亮，则记录为 1，否则为 0.
- ⑤.比对两张图片的指纹，获得汉明距离即可。

对颜色特征的表达方式有许多种, 我们采用直方图进行特征描述。常见的直方图有两种：统计直方图，累积直方图。

4、histogram.py——利用计算直方图实现识别相似图片

- ①.读入图片，并将每个通道的图片值转换成一维数组
- ②.生成并显示每个通道的直方图
- ③.计算每个通道直方图的重合程度

5、histogram_split.py——利用图片分块计算直方图实现识别相似图片

- ①.读入图片，将图片等分为 16 个图片色块
- ②.将 16 个色块每个通道的图片值转换成一维数组
- ③.生成并显示 16 个色块每个通道的直方图
- ④.计算 16 个色块每个通道直方图的重合程度
- ⑤.综合比较结果

五、课程设计创新点说明

基于深度学习的级联相似度排序, 使用预训练好的模型 VGG16/ inceptionV3/ ResNet50 : 识别出图片中车辆的类型(modelID); 识别 (Color identification) 出图片中车辆的颜色(colorID); (x)Hash&(y)Histogram: 计算出汉明距离(Hamming Distance)或直方图的重合程度并按序检索出结果(x)Hash, 包括: aHash, dHash, pHash, (y)Histogram: Histogram, Histogram_split。即先通过对 val(query) 进行颜色和车型的预测排序, 再由高到低对 train(ref) 进行相似度的计算, 可快速精准搜索出相似图片。

六、实验结果与分析

对车辆精准识别的结果，项目使用 MAP@K (K=200) 的评测指标对其进行评估。具体的评估方法可以分为三步：P——AP@K——MAP@K，得到的 MAP 指标值在 0 到 1 之间，值越大代表准确率越高。

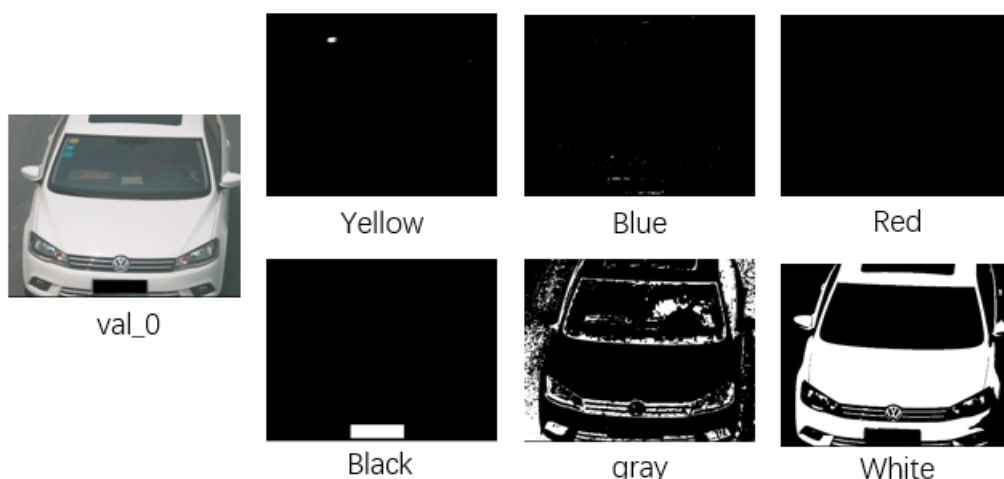
在实际操作中，由于测试集的最终识别结果我们不得而知，因此实际的 MAP@K 指标值我们也无法计算，所以在下面的实验结果分析中，我们将从下面的三个方面分析算法的结果。

（一） 车辆型号识别模型性能效果

如前所述，车辆型号识别的模型我们使用了基于迁移学习的深度卷积神经网络模型 (VGG16)，该模型在 ImageNet 上的分类结果可以达到 92.5% 的 top-5 准确率。在对 VGG16 进行迁移学习的过程中，我们将模型参数设置为[epochs = 60, batch_size = 100, learning-rate=1e-4]，并将优化函数设置为交叉熵，最终达到的在训练集上的型号分类准确率达到 97%，可见这是一个相当优秀的分类效果。

（二） 车身颜色识别算法性能效果

对车身颜色的识别算法性能效果的分析，我们可以从下面的实际例子来看。



在颜色的识别算法中，我们统计不同颜色区域面积，如果使用某种颜色分量过滤背景后，出现车辆的轮廓且可视面积最大，则可认为车辆为该颜色，由上图可以直观看出 val_0.jpg 的颜色为白色且与真实情况相符。将多张图片进行颜色识别，发现识别效果较好，因此构建基于 OpenCV 的 HSV 颜色空间的模型进行识别。

（三） Sample 的结果展示

在性能分析的阶段，我们对某一次查询的结果进行了分析和验证。算法的输入图片如下图左所示。我们对算法的输出结果随机抽选了几张图片，如图右所示。



从算法的输出结果来看，我们可以发现右边 4 辆车与左边查询的车辆在外形上几乎完全相同，但值得注意的是它们并不是同一辆车，因为车辆的品牌显然是不同的。造成这种问题的原因是因为我们的模型只考虑了车辆的型号和颜色而没有考虑车辆的品牌，但这个问题在当前的项目中其实是可以容忍的，因为就数据集来说，与待查询车辆完全相同的车辆图片并不是全部都有 K 张，所以最终的结果必然会包含与该车相似的其他车辆的图片，就如同这个例子显示的那样。进一步的优化方法是同时考虑车辆的品牌信息，但这个需要数据集作相应的更新。

七、课程总结与收获

一个学期很快就过去了，课程期间认真学习了田永鸿老师讲授的《机器学习原理》这门课程，在课堂上学习了在以不同的图像视频分析处理任务中常用的机器学习理论、方法与技术：包括 Bayes 决策与分类、核方法与支持向量机、深度学习等，完成了 paper reading 和课程作业以及大作业；同时自己也在课余翻阅了周志华《机器学习》等书籍。通过对机器学习基础理论的学习，使得自己能够利用机器学习来解决图像视频分析中的挑战性任务，为今后的相关研究与开发打下基础。在此，深深的感谢田老师对自己的教导和助教对自己的帮助！

——杨国峰

《机器学习》课是我接触的第一门机器学习和 AI 相关的课程，在这门课的学习过程中，我经过了入门到迷惑再到豁然开朗的过程，这个过程对我是非常有益的，它帮助我建立起对整个领域的认识，加深了很多概念的理解，并自己实际上手完成了第一个机器学习项目。虽然项目的实际难度并不大，但第一步总是最困难的，万幸我在这门课中学习到很多有用又实用的东西，我非常开心。同时还要感谢田老师和助教的悉心讲课和答疑解惑，祝你们生活愉快。

——张越