Human Eye Project Report - Stage One

张晨阳 郭丹琪 信息学院

2020 年 10 月 30 日

1 Introduction

基本项目任务：收集尽可能多的视觉数据，包括图像或视频，构建识别模型，从视觉内容中生成文本信息。

我们希望实现的扩展任务：在能识别出图片上物体类别的基础上，在类别内部进行更细致的识别和分辨。在本项目中，我们希望提供一个植物图像识别应用 PLANET(Plant Expert)，根据植物的图片，识别植物种类，帮助人们快速便捷地辨别植物。

接下来，我们将分三部分介绍第一阶段的项目提案与进展。首先，我们在第二部分介绍第一阶段的主要工作内容。然后，我们在第三部分具体分析我们的扩展任务——PLANET。最后，我们在第四部分提出下一阶段的工作计划。

2 What We Do in Stage One

在这一部分，我们主要介绍第一阶段完成的工作。我们首先对现有相关工作进行了调研，然后我们提出了多种项目选择并最终确定了项目 PLANET，最后我们学习了相关工作的识别模型和开发思路。

2.1 Research

从图像中提取信息首先要做的是目标检测，即从图像中检测并定位到待识别的物体。我们对当前已有的目标检测方法进行了调研。传统目标检测系统采用 Deformable Parts Models (DPM)，通过滑动框方法提出目标区域，然后采用分类器来实现识别。此外，还有 R-CNN 系列算法，例如 RCNN、Fast RCNN 和 Faster RCNN，这些算法采用的是 Region Proposal Methods。首先生成潜在的 bounding boxes，然后采用分类器识别这些 bounding boxes 区域。最后通过 post-processing 去除重复 bounding boxes 来进行优化。这类方法流程复杂，存在速度慢和训练困难的问题。最后我们发现了 YOLO 算法，它将目标检测问题转换为直接从图像中提取 bounding boxes 和类别概率的单个回归问题，只需一眼（you only look once，YOLO）即可检测目标类别和位置，预测流程简单，速度快。

YOLO 采用单个卷积神经网络来预测多个 bounding boxes 和类别概率，与滑动窗口方法和 Region Proposal Methods 不同，YOLO 在训练和预测过程中可以利用全图信息，正确率较高。在调研过后，我们决定采用 YOLO 算法，并决定对其进行学习。

2.2 Project Selection

在对图像识别算法、目前已有的图像识别应用进行了调研后，我们考虑在基础的 human eye 的任务上进行扩展，做出有实际应用价值的功能。

我们首先想出了一些应用，主要分为两类。第一类为读取图像中的文字内容，具体应用如下：

- 搜题软件中提取图片中的题目文字。
- 提取图片中的文字并转换为文本信息。
- 提取视频中的英文字幕并实时翻译为中文。

第二类为读取图像中的物体信息，具体应用如下：

- 识别植物图像，告知用户植物种类。
- 拍照识别物体，教小孩子认识物品。
- 读取人像照片，识别人物的年龄、性别、情绪等。
- 对视频做睡意检测：当用户昏昏欲睡时，通过检测眼睛来发出警报，预防事故发生。

我们分别调研了每个项目的创新性与可行性。第一类读取图像中的文字内容，目前已有成熟的开源应用框架，缺少创新性。拍照识别物体这两个方向，目前已有成熟的应用实现。人脸识别情绪、困意检测这两个方向，缺少满足需求的足够大的训练数据集。因此，在结合了对创新性与可行性的具体分析后，我们决定选择植物图像识别这一应用方向，并将我们的项目命名为 PLANET (Plant Expert)。

2.3 Learn Related Works

在调研过后，我们选择了 YOLO [1] 框架，并对其进行了学习。YOLO (You Only Look Once) 是计算机视觉领域中目标检测的端到端的方法。YOLO 框架与 RCNN 系列算法不一样，RCNN 等算法将目标识别任务分为目标区域预测和类别预测等多个流程，而 YOLO 将目标区域预测和目标类别预测整合于单个神经网络模型中，实现在准确率较高的情况下快速目标检测与识别，更加适合现场应用环境。在实现上，YOLO 首先将图像分为多个网格，对每个网格应用图像分类和定位处理，获得预测对象的边界框及其对应的类概率。最终可以在图片中框出目标物，并给出相应名称。模型采用卷积神经网络结

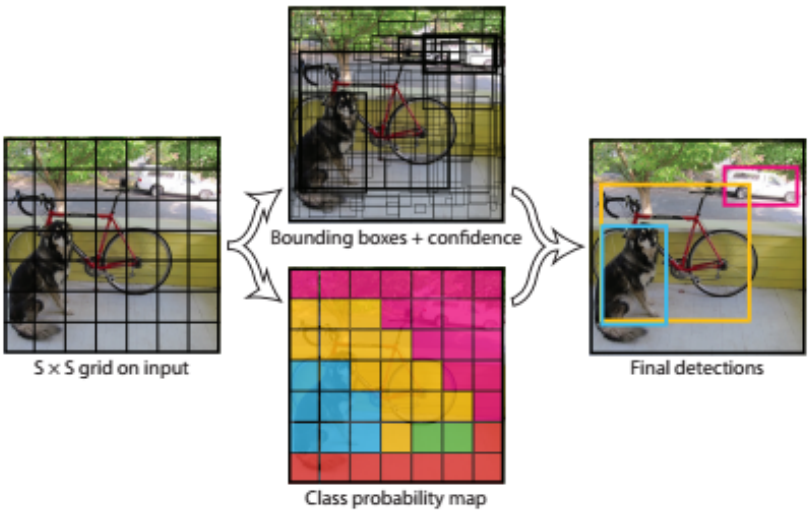

图 1: YOLO 识别过程

构。开始的卷积层提取图像特征，全连接层预测输出概率，模型结构类似于 GoogleNet。同时，模型中采用了非极大值抑制判断预测的边界框的划分结果以提高准确率。

具体来说，YOLO 首先将图像分为 $S \times S$ 的网格。如果一个目标的中心落入网格，该网格就负责检测该目标。每一个网格预测 bounding boxes 和该 boxes 的置信值 confidence。置信值代表 box 包含一个目标的置信度，如果没有目标，则置信值为零。定义置信值为 $Pr(Object) \times IOU_{pred}^{truth}$。其中 $Pr(Object)$ 是该 box 里包含目标的概率，IOU 是非极大值抑制方法中判断预测的边界框的划分结果的值，IOU = 实际边框界与预测边框界的交叉面积/联合的面积。

每一个 bounding box 包含 5 个值：x，y，w，h 和 confidence。(x，y) 代表与格子相关的 box 的中心。(w，h) 为与全图信息相关的 box 的宽和高。每个网格的预测条件概率值 C 为 $Pr(Class_i|Object)$。概率值 C 代表了网格包含一个目标的概率，每一网格只预测一类概率。在测试时，每个 box 通过类别概率和 box 置信度相乘来得到特定类别置信分数：$Pr(Class_i|Object) \times Pr(Object) \times IOU_{pred}^{truth} = Pr(Class_i) \times IOU_{pred}^{truth}$。这个分数代表该类别出现在 box 中的概率和 box 和目标的合适度。在利用上述方法得到每一个目标的 bounding box 后，就可以对 box 中的相应目标进行检测。上述内容在图 1中展示。

整个模型采用卷积神经网络结构。开始的卷积层提取图像特征，全连接层预测输出概率，模型结构类似于 GoogleNet，在图 2和图 3中展示。

我们安装并配置了 YOLO，进行了两类测试。首先我们用训练好的模型 *yolov3.weights*，分别对 YOLO 的例图与我们选取的图片进行了识别。识别结果分别展示在图 4和图 5中。此外，我们用已经训练好的人脸识别模型 *Yolov3-tiny-Face-weights* 测试了 YOLO 的人脸识别效果，结果在图 6中展示。

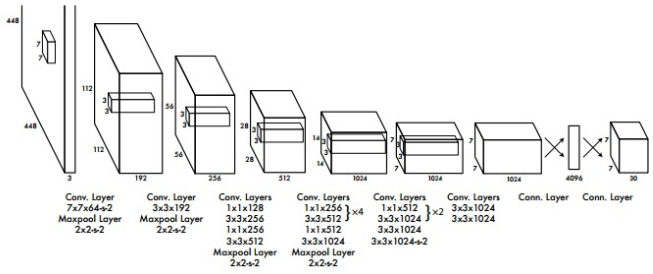

图 2: YOLO 神经网络结构

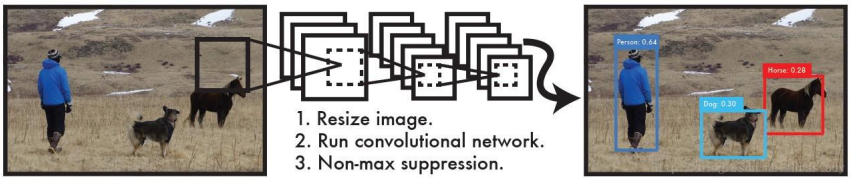

图 3: YOLO 神经网络结构 2

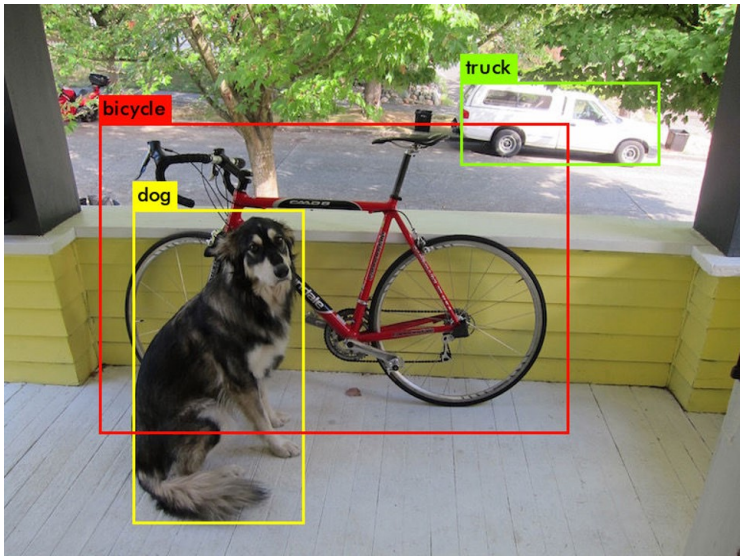

图 4: YOLO 例图的 YOLO 识别

图 5: 选取图的 YOLO 识别

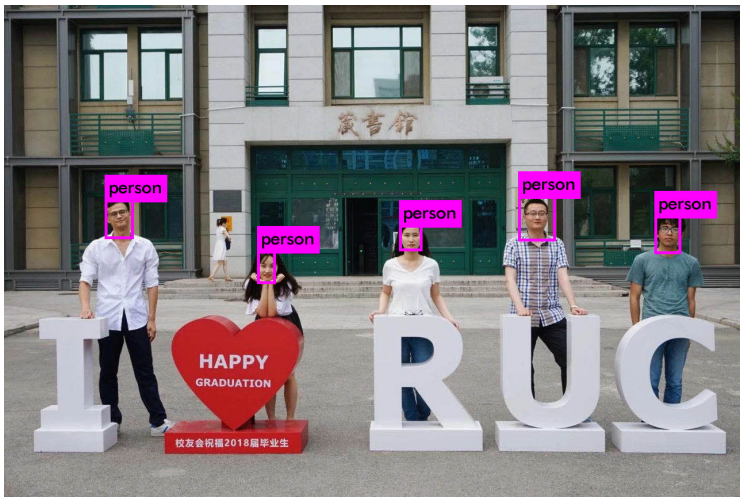

图 6: 通过 YOLO 实现的人脸识别

3 Analysis of Our Project

3.1 Contribution

在这一部分，我们概括工作的贡献。

- 我们提出了首个开源的植物图像识别工具，帮助人快速辨别植物。
- 我们探究了已有图像识别框架在识别相似植物上的效果，揭示了这些框架识别并分辨相似物体的能力。

3.2 Motivation

在这一部分，我们阐述项目开发的动机。

- 当人们见到有趣的植物，想要知道它的品种时，只能通过查阅植物百科的方式。这种方式对于不随身携带植物百科书的人来说，十分不方便，且需要人为翻书搜索，耗时很长。现在没有开源图像识别框架能准确识别植物。因此我们希望开发一个开源植物图片识别框架，来填补这个缺口，帮助人们快速、准确地识别并分辨植物。
- 现有的图像识别框架大部分是很一般、概括性的识别，例如 YOLO 可以识别出人和狗，但不能分辨狗的种类是藏獒还是吉娃娃。我们希望进一步探究这些图像识别框架的能力。在本项目中，我们探究这些框架在某个特定小领域上，给近似图像分类的准确度。更具体的，我们探究现有框架在植物领域，分辨植物类别的准确度。这样的探究帮助我们更好地了解现有图像识别框架的能力和限制，为后续科研工作者的研究与模型改进提供思路。

3.3 Challenge

在这一部分，我们详细阐述工作的挑战性。

Challenge 1: 寻找合适的、大量的数据集。 图像试别问题一般采用监督学习算法解决。识别模型在使用前需要进行训练，训练数据集的准确率、规模对识别模型的准确率有很大的影响。目前已有的开放数据集质量参差不齐，各数据集的组织方式、标签表示、侧重点不同。因此在这个问题中，选择并整合出适合识别模型、规模大、覆盖范围广的数据集是一件很重要且有挑战性的事情。

Challenge 2: 探索已有工作使用的模型与开发思路。 在第二部分中，我们给出了对已有识别框架的调研。这些已有研究采用了多种神经网络来进行图像识别。对于没有系统学习过深度学习与神经网络的我们，学习已有工作的神经网络模型和开发思路是一件有挑战性的事情。这些工作不只是应用了普通的神经网络，而是在神经网络的基础上做了

一系列的特异性优化。我们试图学习他们改进神经网络的思路，并将此应用到我们的项目中。

Challenge 3：对识别模型做针对性优化。 我们要实现的识别模型是分辨出各种植物的种类，这一模型仅针对于植物领域，有一定的特异性；对识别精度有很高的要求，需要区分出相似的植物。由于我们模型的特异性，直接采用一般的识别模型不能达到最好的识别分辨效果。因此我们需要在已有识别模型的基础上，做针对本项目的优化。这就需要我们深入理解已有识别模型的设计思路，即挑战 2 阐述的内容，然后在充分了解现有工作的基础上，分析我们项目的特点，做针对性优化。

4 Future Plan

在这一部分，我们阐述下一阶段的计划。

- 继续研究 YOLO 及其他图像识别框架的原理、开发思路。
- 在现有框架的基础上，做针对 PLANET 的设计和优化。
- 寻找并构建合适的训练数据集。
- 测试训练好的模型的识别效果，并根据效果不断优化模型。

Reference

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.