**2019基于人工智能的视觉识别**

**技术项目**

**详细方案**

团队：窝窝头一块钱四个小组

目录

[1. 前言 1](#_Toc41400147)

[1.1 编写目的 1](#_Toc41400148)

[1.2 背景 1](#_Toc41400149)

[1.3 术语 2](#_Toc41400150)

[1.4 参考/引用资料 2](#_Toc41400151)

[2. 项目概述 3](#_Toc41400152)

[2.1 适用范围及系统特性简要说明 3](#_Toc41400153)

[2.2 项目创意及特色 3](#_Toc41400154)

[2.2.1 项目创意 3](#_Toc41400155)

[2.2.2 项目特色 4](#_Toc41400156)

[2.3 功能简介 4](#_Toc41400157)

[2.4 开发工具与技术 5](#_Toc41400158)

[3. 项目主要功能/流程的详细介绍​ 5](#_Toc41400159)

[3.1 实验数据制作 6](#_Toc41400160)

[3.1.1 数据获取 6](#_Toc41400161)

[3.1.2数据清洗 6](#_Toc41400162)

[3.1.3 数据标注 7](#_Toc41400163)

[3.2 YOLOv3原理 7](#_Toc41400164)

[3.2.1 特征提取网络Darknet-53 7](#_Toc41400165)

[3.2.2 YOLO v3设计思想 8](#_Toc41400166)

[3.3 改进方法 9](#_Toc41400167)

[3.3.1基于图像金字塔的多尺度特征检 9](#_Toc41400168)

[3.3.2 k-means维度聚类算法 11](#_Toc41400169)

[3.3.3 多尺度训练 12](#_Toc41400170)

[4. 重点功能函数说明​ 13](#_Toc41400171)

[5. 市场分析及行业分析 15](#_Toc41400172)

[5.1 市场分析 15](#_Toc41400173)

[5.1.1 政治因素 15](#_Toc41400174)

[5.1.2 经济因素 16](#_Toc41400175)

[5.1.3 社会因素 16](#_Toc41400176)

[5.1.4 技术因素 17](#_Toc41400177)

[5.2 行业分析 17](#_Toc41400178)

[6. 风险分析与对策 19](#_Toc41400179)

[6.1 风险分析 19](#_Toc41400180)

[6.1.1 技术风险 19](#_Toc41400181)

[6.1.2 市场风险 19](#_Toc41400182)

[6.1.3 管理风险 20](#_Toc41400183)

[6.1.4 安全风险 20](#_Toc41400184)

[6.2 风险应对对策 20](#_Toc41400185)

[6.2.1 风险应对原则 20](#_Toc41400186)

[6.2.2 技术风险对策 21](#_Toc41400187)

[6.2.3 市场风险对策 21](#_Toc41400188)

[6.2.4 管理风险对策 21](#_Toc41400189)

[6.2.5 安全风险对策 22](#_Toc41400190)

[7. 结语 22](#_Toc41400191)

# 前言

## 1.1 编写目的

本说明书给出基于人工智能的视觉识别技术项目的设计说明，包括最终实现的项目必须满足的功能、效率、采用实现技术的详细说明。

**目的在于：**

* 为编码人员提供依据；
* 为修改、维护提供条件；
* 项目负责人将按此计划书的要求布置和控制开发工作全过程。

**本说明书的预期读者包括：**

* 项目开发人员，特别是编码人员；
* 软件维护人员；
* 技术管理人员；
* 项目负责人和全体干系人。

## 1.2 背景

智能识别一直是人工智能领域研究的重点问题，它被广泛地应用到生活的各个方面。伴随着经济的发展，工厂的数量日益增加，同时工人因在生产车间不佩戴安全帽而引发的安全事故也层出不穷，所以需要加大和完善监管体制。如果人为监管难免费时费力，存在疏忽。本项目旨在实现一款神经网络软件，做到对生产车间未佩戴安全帽的行为做到实时识别，实时监测，实时报警。

待开发项目的名称：基于人工智能的视觉识别技术项目

此项目任务提出者：文思海辉技术有限公司

此项目任务开发者：窝窝头一块钱四个

此项目任务对象：测试图片

## 1.3 术语

* **数据爬取：**通过爬取多个不同的关键词，爬取百度图片和Google图片。
* **数据清洗：**爬取得到的图片包含大量重复的，或者是并不包含ROI的图片，需要过滤掉大量的这些图片。
* **数据标注：**对图像识别目标进行标记，并生成相对应的配置文件。
* **目标检测：**找出图像中所有感兴趣的物体，包含物体定位和物体分类两个子任务，同时确定物体的类别和位置。

## 1.4 参考/引用资料

* YOLO: Real-Time Object Detection:
* (https://blog.csdn.net/haishu\_zheng/article/details/80430106)
* PyTorch-YOLOv3:
* (<https://github.com/eriklindernoren/PyTorch-YOLOv3>)
* Implement YOLO v3 from scratch： (<https://blog.paperspace.com/how-to-implement-a-yolo-object-detector-in-pytorch/>)
* A PyTorch implementation of a YOLO v3 Object Detector：（<https://github.com/ayooshkathuria/pytorch-yolo-v3>）
* BVS安全帽识别系统（安全帽佩戴检测）：
* (<https://cloud.tencent.com/developer/article/1131994>)
* A Y0LOv3-based non-helmet-use detection for seafarer safety aboard merchant ships:
* （<https://iopscience.iop.org/article/10.1088/1742-6596/1325/1/012096/pdf> ）
* 需求规格说明标准规范:
* （http://blog.sina.com.cn/s/blog\_4902a6390102w1k9.html）
* 一文盘点2012年以来国内大数据相关政策:  
  (https://blog.csdn.net/enohtzvqijxo00atz3y8/article/details/80730754）
* Rubaiyat, A. H., Toma, T. T., Kalantari-Khandani, M., Rahman, S. A., Chen, L., Ye, Y., & Pan, C. S. (2016, October). Automatic detection of helmet uses for construction safety. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW)* (pp. 135-142). IEEE.
* RongXin, W. (2019, November). Research on safety helmet wearing YOLO-V3 detection technology improvement in mine environment. In *Journal of Physics: Conference Series* (Vol. 1345, No. 4, p. 042045). IOP Publishing.
* Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint* arXiv:1804.02767.

# 项目概述

## 2.1 适用范围及系统特性简要说明

本项目的应用对象主要是工业生产的各个生产车间，用深度学习技术检测人员是否佩戴安全帽。随着社会的快速发展，在一些生产和作业场地安全事故频频发生。大部分的事故原因是工人未遵守规章制度，不正确佩戴安全帽，因此本项目是通过实时检测职工是否佩戴安全帽，以此来保障生产车间的生产安全。

## 2.2 项目创意及特色

### 2.2.1 项目创意

本项目是基于TensorFlow和YOLOv3的深度学习框架实现生产环境安全帽的实时检测。TensorFlow作为“第二代机器学习系统”，相比于其他，具有很好的架构灵活性和可移植性。YOLOv3是一个庞大而丰富的深度卷积神经网络模型，它的快速检测和识别的精度，能够非常完美地实现本项目基于场景的实时检测的需求。在大数据飞速发展的今天，基于YOLOv3的实时检测技术已经成为了主流，它被广泛地应用于生产生活的各个领域。因此，在未来的大趋势下，随着行业的不断发展以及需求的不断细分，利用TensorFlow和YOLOv3的深度学习模型将会对安全帽识别系统有进一步的相关功能的优化，从而为企业带来更加便捷可靠的安全管理。

### 2.2.2 项目特色

1：采用争对大量数据集的深度学习框架从而可以自动地构建的数据集上归纳出一套分类规则。

2：采用目前业界普遍认可的TensorFlow和YOLOv3进行图片识别和图片处理。

3：采用多目标跟踪方法：使用Deep\_SORT将唯一的ID分配给检测到的人，并通过视频的连续帧对其进行跟踪。

4：提供实时识别，每秒钟帧数 (fps) 达到35帧。

5：提供警示功能，识别出的用户超过5秒没有佩戴安全帽，系统发出警示。

## 2.3 功能简介

该项目旨在利用现有的闭路电视摄像机基础设施，通过提供实时警报，帮助主管有效监控工人是否佩戴安全帽。

本项目使用三个开源的检测数据集，一部分来自开源的安全帽检测数据集(SafetyHelmetWearing-Dataset,SHWD)，总共有7581张图像，包含9044个佩戴安全帽的bounding box（正类），以及111514个未佩戴安全帽的bounding box(负类)，所有的图像用labelimg标注出目标区域及类别。其中每个bounding box的标签：“hat”表示佩戴安全帽，“person”表示普通未佩戴的行人头部区域的bounding box。一部分来自Stanford 40 Actions，另外本数据集中person标签的数据大多数来源于SCUT-HEAD数据集，用于判断佩戴安全帽的人。

在对已知数据集进行清洗（去掉重复的，或者是并不包含ROI的图片），然后进行bounding box标注。程序自动将数据集按0.8比率进行随机切割，0.8为训练集，0.2为训练集。使用YOLO v3官网上提供的权重参数作为网络训练的初始化参数，随机使用自制安全帽佩戴检测训练数据集中的图片进行网络参数微调（finetune），使得整个网络检测效果达到最优。采用YOLOv3识别技术检测安全帽的程序流程图如图1所示。

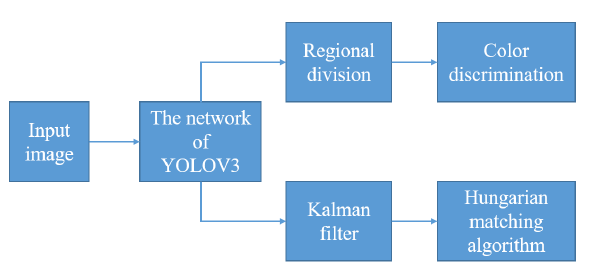


图1 程序流程图

## 2.4 开发工具与技术

|  |  |  |
| --- | --- | --- |
| 分类 | 名称 | 版本 |
| 开发工具 | pycharm |  |
| git |  |
| 重要库 | pandas | 1.0.3 |
| TensorFlow | 1.15.0 |
| numpy | 1.18.1 |
| tqdm | 4.46.0 |
| YOLOv3 |  |
| 代码托管平台 | GitHub |  |

# [3. 项目主要功能/流程的详细介绍​](#_Toc2560_WPSOffice_Level1 )

本项目基于YOLOv3模型为主体，首先采用图像金字塔结构，将不同层次特征进行融合，获取不同尺度的特征图，用于位置和类别预测；然后对目标框维度进行聚类，增加先验框（anchorbox）个数，使得模型能够获取更多的物体边缘信息；最后，在训练过程中，采用多尺寸图片进行训练，使得模型能够适应不同分辨率的图片。实验表明，改进的YOLOv3算法能够在保证检测精度的同时，提高检测速率。

## 3.1 实验数据制作

### 3.1.1 数据获取

本项目使用的数据集来源于开源数据集和网络爬取获得的数据集。深度学习在训练数据的时候很重要的一部分就是数据，因此可以通过爬虫爬取一些相关数据。

#### 3.1.1.1 开源数据集

本项目采用了三个开源的检测数据集：

1. 一部分来自企业开源的安全帽检测数据集(SafetyHelmetWearing-Dataset,SHWD)，总共有7581张图像，包含9044个佩戴安全帽的bounding box（正类），以及111514个未佩戴安全帽的bounding box(负类)，所有的图像用labelimg标注出目标区域及类别。其中每个bounding box的标签：“hat”表示佩戴安全帽，“person”表示普通未佩戴的行人头部区域的bounding box。
2. 一部分来自Stanford 40 Actions
3. 本数据集中person标签的数据大多数来源于SCUT-HEAD数据集，用于判断佩戴安全帽的人。

#### 3.1.1.2 爬取网络数据

用Python来爬取相关的安全帽百度图片和Google图片，百度图片用自己编写的访问web页面的方式，通过不同的关键词来多线程爬取相关安全帽的图片，Google图使用google-images-download的方式爬取图片。

### 3.1.2数据清洗

数据清洗是数据准备的重要一步，网络爬取得到的图片具有大量的重复性，或者是有些并不包含ROI的图片，需要过滤掉大量的这些图片，以提高数据质量，过滤过程如下所述：

(1)用已有的行人检测方法过滤掉大部分非ROI图像；

(2)可以使用深度学习模型zoo，比如ImageNet分类预训练好的模型提取特征，判断图像相似度，去除极为相似的图像；

(3)剩余的部分存在重名或者文件大小一致的图像，通常情况下这些都是不同链接下的相同图片，在数量少的情况下可以手动清洗。

### 3.1.3 数据标注

数据标注是数据加工人员通过借助开源标注工具对图像进行批注，是对深度学习数据进行加工的一种行为。图像标注和视频标注按照数据标注的工作内容来分类的话，统称为图像标注。对图像的数据标注是一个相对复杂的过程，标注人员需要对不同的目标标记物用不同的颜色进行轮廓标记，然后对相应的轮廓打标签，用标签来概述轮廓内的内容，以便让模型能够识别图像的不同标记物。

本项目采用开源标注工具labelImg来进行标注，利用该标记工具，在人工操作的时候，我们只需要在图片中标记出自定义的各种目标，该工具会自动生成相应的配置文件。我们结合使用更为省力的方法，数据回灌，也就是先用标注好的一部分数据训练出一个粗糙的检测模型，精度虽然不高，不过可以拿来定位出大致的目标区域位置，然后进行手动调整bounding box位置，这样即省时又省力，反复操作可减少工期。另外，在标注的过程中会出不少问题比如由于手抖出现图中小圈的情形,这种情况会导致标注的xml出现bounding box的四个坐标宽或高相等，显然不符合常理，所以需要手动写脚本检查和处理有这种或者其他问题的xml的annotation，比如还有的检测算法不需要什么都没标注的背景图像，可以检测有没有这种“空”类别的数据，甚至是笔误敲错了类别的标签，这些都需要手动写自动化或半自动化的脚本来做纠错处理，这样的工具在标注时应该经常用到。同时，也可以看出，一旦标注项目形成规模，规范的自动化流程将会节省很多资源。

## 3.2 YOLOv3原理

### 3.2.1 特征提取网络Darknet-53

YOLOv3使用新的网络来实现特征提取。从网络结构上看，相比YOLOv2中的Darknet-19网络，添加了残差单元，使用连续的3×3和1×1卷积层，YOLOv3将其扩充为53层并称该网络为Darknet-53。

该网络结构包含53个卷积层以及5个最大池化层，同时，在每一个卷积层后增加了批量归一化操作（batchnormalization）和去除dropout操作，防止出现过拟合现象，该网络比Darknet-19功能强大，同时比ResNet-101或ResNet-152更有效，在ImageNet数据集下的性能测试结果如表1所示。

表1 ImageNet数据集下的性能测试结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Backbone | Top-1% | Top-5% | Bn Ops/109 | BFLOP/s | 识别帧率/(f**·**s-1) |
| Darknet-19 | 74.1 | 91.8 | 7.29 | 1246 | **171** |
| Resnet-101 | 77.1 | 93.7 | 19.70 | 1039 | 53 |
| ResNet-152 | **77.6** | **93.8** | **29.40** | 1090 | 37 |
| DarkNet-53 | 77.2 | 93.8 | 18.70 | **1457** | 78 |

由表1可以看出，Darknet-53比ResNet-101检测效果更好，且速度提高了1.5倍，Darknet-53比ResNet-152具有相似的性能，速度提高了2倍，与Darknet-19相比，检测准确率有较大提升，虽然速度不及Darknet-19，但仍然满足实时性要求。

### 3.2.2 YOLO v3设计思想

YOLOv3是采用多个尺度融合的方式来进行预测。使用类似FPN（FeaturePyramidNetworks）的融合做法，在多个尺度的特征图（featuremap）上进行位置和类别预测，提高了目标检测准确率。在YOLOv3中，同样使用了维度聚类（dimensionclusters）作为先验框来预测边界框。通过k-means方法对数据集中的目标框进行维度聚类，得到9个不同大小的先验框，并将其均分到多个尺度的特征图上，尺度更大的特征图使用更小的先验框，该方案使得最终获得的先验框数量比YOLOv2多，特征提取效果更好。在类别预测中，YOLOv3不使用softmax对每个框进行分类，而是采用多个独立的逻辑（logistic）分类器，在训练过程中，使用二元交叉熵损失（binarycross-entropyloss）来进行类别预测。YOLOv3的网络架构如图2所示。



图2 YOLOv3网络架构图

## 3.3 改进方法

YOLOv3在目标检测领域已经取得了很好的检测效果，但针对YOLOv3所做的物体检测均在常见数据集上实现，而对于特定的安全帽佩戴数据集，需要对YOLOv3进行一些改进，以适应于特定的安全帽佩戴检测任务，其改进方法如图3所示。



图3 改进方法示意图

基于图像金字塔结构，改进YOLOv3算法模型，将不同层次的特征图进行融合连接，得到3组预测特征图，并在这3组预测特征图上进行位置和类别的预测。在自制的安全帽佩戴数据集上使用k-means算法，确定先验框维度。YOLOv3算法中得到的先验框维度是在COCO数据集上训练得到，且将其参数均分到三种不同尺度上，而对于安全帽佩戴检测这一具体任务，需要针对其特定数据集进行聚类操作，得到相应的聚类中心。由于YOLOv3中去除了全连接层，因此在训练过程中，可以随时改变输入尺寸，使得训练后的模型能适应不同尺度的输入图像，其中，输入图像为自制的安全帽佩戴数据集。

### 3.3.1基于图像金字塔的多尺度特征检

YOLOv3中使用Darknet-53网络来提取特征，利用特征可视化技术，能够清楚地看到各个层级的特征效果。图4选取了Darknet-53中的conv1\_1层以及conv53\_1层的特征图形进行分析，可以看出，低层级的特征具有丰富的细节和定位信息，较高层级的特征具有丰富的语义特征，从低层级到高层级，细节信息不断减少，而语义信息不断增加，对于位置预测而言，需要更多的低层特征信息，对于类别预测而言，需要更多高层的局部信息，因此，可以考虑基于图像金字塔的模式，使用上采样将高层特征与低层特征进行融合，获取不同尺度的特征图，用于位置和类别预测。



图4 Darknet-53特征可视化

图5中右边的特征金字塔是由左边的特征金字塔生成的，整个过程为：首先对输入图像进行深度卷积，然后对layer2上面的特征进行卷积操作，对layer4中的特征进行上采样操作，使其具有相同的尺寸，然后对处理后的layer2和layer4执行卷积和操作，将获得的结果输入到layer5层中。同理，相应地对多个层之间进行特征融合，得到用于预测的多组特征图。基于此方案，将处理过的低层特征和高层特征进行累加，这样做的目的是，由于低层特征可以提供更加准确的位置信息，而多次下采样和上采样操作使得深层网络的定位信息存在误差，因此将其结合起来使用，构建了一个更深的特征金字塔，融合了多层特征信息，并在不同的特征图上进行预测。



图5 特征融合金字塔

借鉴上述特征融合思路，对YOLOv3算法进行改进，使用上采样将高层特征与低层特征进行融合，最终得到3组特征图，并利用这3组不同尺度的特征图进行预测。改进的网络结构如图6所示。具体网络结构改进细节：首先通过Darknet-53得到特征金字塔，将conv53层进行连续的1×1和3×3卷积操作，得到一组待处理yolo层，然后将该层进行一组1×1和3×3卷积操作，得到小尺度yolo层；同时，将该层进行上采样，与Darknet-53中的conv45层进行卷积和操作，同样使用连续1×1和3×3卷积操作，得到第二组待处理yolo层，将该层进行一组1×1和3×3卷积操作，得到中尺度yolo层；同时，将该层进行上采样，与Darknet-53中的conv29层进行卷积和操作，同样使用连续1×1和3×3卷积操作，得到第三组待处理yolo层，将该层进行一组1×1和3×3卷积操作，得到大尺度yolo层。经过上述操作，得到3组不同尺度的yolo特征层，并使用这3组特征层进行位置和类别预测。



图6 改进的网络结构图

### 3.3.2 k-means维度聚类算法

YOLOv3借鉴了YOLOv2中使用先验框来预测边界框坐标的方法，不同的是，YOLOv3使用k-means算法预测得到9个先验框，并将其均分到3个尺度特征图，其中，尺度更大的特征图使用更小的先验框，以此来获得更多的目标边缘信息。

YOLOv3中计算得到的9组先验框维度分别为（10，13），（16，30），（33，23），（30，61），（62，45），（59，119），（116，90），（156，198），（373，326），但是在实际安全帽佩戴检测任务中，YOLOv3算法中计算得到的先验框维度不适合安全帽佩戴检测场景。待检测目标框尺寸主要分布如图7所示，因此使用YOLOv3算法中原有的先验框维度很难得到精确的目标框信息。



图7 待检测目标框的尺寸

因此，在安全帽佩戴检测场景下，使用k-means算法对自制的安全帽佩戴数据集进行聚类分析，得到9组先验框维度中心，分别为：（59，22），（68，30），（75，35），（88，37），（90，40），（99，45），（108，54），（119，57），（128，63），将其面积从小到大排列，均分到3个不同尺度的特征图上，其中，尺度更大的特征图使用更小的先验框。最后，将用此聚类中心进行安全帽佩戴检测实验。

### 3.3.3 多尺度训练

在YOLO检测算法中采用卷积网络来提取特征，然后使用全连接层来得到预测值。但由于全连接层的存在，网络在训练过程中输入图片的尺寸必须固定，因此最终训练得到的网络对不同尺寸的测试图片不具备鲁棒性。针对自制的安全帽数据集，其输入图像具有不同大小的尺寸，因此，为了增强模型对图像尺寸的鲁棒性，可以采取多尺度训练的策略。具体来说，在YOLOv3网络中移除全连接层，改为全卷积操作。图8显示了将全连接层转化为卷积层的过程。



图8 全连接层转化为卷积层的过程

# [4. 重点功能函数说明​​](" \l "_Toc29107_WPSOffice_Level1 )

4.1 def predict\_transform(prediction, inp\_dim, anchors, num\_classes, CUDA = True)

在特征图上进行多尺度预测, 在GRID每个位置都有三个不同尺度的锚点。predict\_transform()利用一个scale得到的feature map预测得到的每个anchor的属性(x,y,w,h,s,s\_cls1,s\_cls2...)

维度为(batch\_size, num\_anchors\*bbox\_attrs, grid\_size, grid\_size)，维度变换成(batch\_size,grid\_size\*grid\_size\*num\_anchors, 5+类别数量)的tensor，得到每个方框在网络输入图片(416x416)坐标系下的(x,y,w,h)以及方框含有目标的得分以及每个类的得分。

batch\_size = prediction.size(0)

stride表示的是整个网络的步长，等于图像原始尺寸与yolo层输入的feature mapr尺寸相除，因为输入图像是正方形，所以用高相除即可

stride =  inp\_dim // prediction.size(2)#416//13=32

feature map每条边格子的数量，416//32=13

grid\_size = inp\_dim // stride

一个方框属性个数，等于5+类别数量

bbox\_attrs = 5 + num\_classes

anchors数量

num\_anchors = len(anchors)

4.2 def letterbox\_image(img, inp\_dim):

lteerbox\_image()将图片按照纵横比进行缩放，将空白部分用(128,128,128)填充,调整图像尺寸

具体而言,此时某个边正好可以等于目标长度,另一边小于等于目标长度，缩放后的数据拷贝到画布中心

取min(w/img\_w, h/img\_h)这个比例来缩放，缩放后的尺寸为new\_w, new\_h,即保证较长的边缩放后正好等于目标长度(需要的尺寸)

new\_w = int(img\_w \* min(w/img\_w, h/img\_h))

new\_h = int(img\_h \* min(w/img\_w, h/img\_h))

resized\_image = cv2.resize(img, (new\_w,new\_h), interpolation = cv2.INTER\_CUBIC) #将图片按照纵横比不变来缩放为new\_w x new\_h，768 x 576的图片缩放成416x312.,用了双三次插值

创建一个画布, 将resized\_image数据拷贝到画布中心。

canvas = np.full((inp\_dim[1], inp\_dim[0], 3), 128)

4.3 def arg\_parse():

视频检测模块的参数转换

创建一个ArgumentParser对象，格式: 参数名, 目标参数(dest是字典的key),帮助信息,默认值,类型

parser.add\_argument("--bs", dest = "bs", help = "Batch size，默认为 1", default = 1)

parser.add\_argument("--confidence", dest = "confidence", help = "目标检测结果置信度阈值", default = 0.5)

parser.add\_argument("--nms\_thresh", dest = "nms\_thresh", help = "NMS非极大值抑制阈值", default = 0.4)

if ret:

将图片按照比例缩放缩放，并且将HxWxC转换为CxHxW

img = prep\_image(frame, inp\_dim)

cv2.imshow("a", frame)

得到图片的W,H,是一个二元素tuple.因为我们不必再处理 batch，而是一次只处理一张图像，所以很多地方的代码都进行了简化。

因为一次只处理一帧，故使用一个元组im\_dim替代 im\_dim\_list 的张量。

im\_dim = frame.shape[1], frame.shape[0]

先将im\_dim变成长度为2的一维行tensor，再在1维度(列这个维度)上复制一次，变成1x4的二维行tensor[W,H,W,H]，展开成1x4主要是在后面计算x1,y1,x2,y2各自对应的缩放系数时好对应上。

im\_dim = torch.FloatTensor(im\_dim).repeat(1,2)

repeat()可能会改变tensor的维度。

# 市场分析及行业分析

## 5.1 市场分析

### 5.1.1 政治因素

为了深入贯彻落实党的十九大精神，全力打造安全生产治理新格局，安全问题得到了众多企业的广泛关注，开始着手于大力宣传安全问题的重要性。在一些户外的作业以及企业的相关工程的实施中，安全永远排在众多问题的首位。在2017年3月，国务院在《安全生产标准“十三五”发展规划》规章中也明确提出：“深入推进施工现场的安全生产改革发展，加快提升从业人员的安全意识，规范安全行为，满足群众安全需求”等。不管是在生产车间，还是在建筑工程建设、化工厂、航运以及电力工程等领域中，安全都是工人最重要亦是最基本的一个需求。



图9 2011年至2017年第三季度死亡人数和已发生事故起数量数据

例如，据住建部网站发布的房屋市政工程生产安全事故情况每月通报情况，统计2011年至2017年第三季度截止已发生事故起数量数据如图9所示，每年死亡人数都在五百人以上。同时表明行业发展速度与工人安全需求的冲突日渐增加，采用有效的图像识别技术自动检测工人是否佩戴安全帽是非常有必要的，我们采用先进的YOLOv3技术可以实现对生产环境安全帽的实时检测。

### 5.1.2 经济因素

在变电站、建筑场所、化工厂等一些安全生产规范中，明确指出不允许未佩戴安全帽进入作业场所。尽管各单位经常进行安全教育，但是并不能保证人员时刻佩戴安全帽。现在很多地方还是采取人力盯梢的方式，用人眼通过相机或者现场检查有无场内人员不佩戴安全帽的情况，但巡查人员不可能全天候盯着，需要借助科技手段来实现监督工人是否佩戴安全帽。利用 YOLO 算法识别生产车间工人是否佩戴安全帽问题可做到全天候无作息实时工作、无需人工操作、工作效率极高，即省时又省力。

### 5.1.3 社会因素

识别工作人员是否佩戴安全帽是施工现场安全管理的一项基础性且重要性的工作。随着计算机的硬件与软件技术的发展给计算机视觉方法带来研究上质的提升，利用YOLO v3算法的计算机视觉方法可以识别出目标类别与目标对象坐标概率。在车间安全帽检测系统中，采用YOLO v3目标检测技术可进行实时识别工人的安全帽佩戴情况，将监控系统的图像或视频流训练该算法参数，训练成功后可对图像或视频流进行实时识别，并提供了良好的智能化监测方式。通过对施工作业厂区进行这样的实时的监督，为工人敲响安全的警钟，在提高工人的安全意识的同时减少安全事故发生。基于TensorFlow和YOLOv3的深度学习模型也将会带来很好的商业价值和社会应用价值。

### 5.1.4 技术因素

目标检测模型的主要性能指标是检测准确度和速度，其中准确度主要考虑物体的定位以及分类准确度。目前，主流的目标检测算法大概可分成两大类：

1. One-Stage目标检测算法，这类检测算法不需要Region Proposal阶段，可以通过一个Stage直接产生物体的类别概率和位置坐标值，比较典型的算法有YOLO、SSD和CornerNet；
2. Two-Stage目标检测算法，这类检测算法将检测问题划分为两个阶段，第一个阶段首先产生候选区域（Region Proposals），包含目标大概的位置信息，然后第二个阶段对候选区域进行分类和位置精修，这类算法的典型代表有RCNN，Fast RCNN，Faster RCNN等。

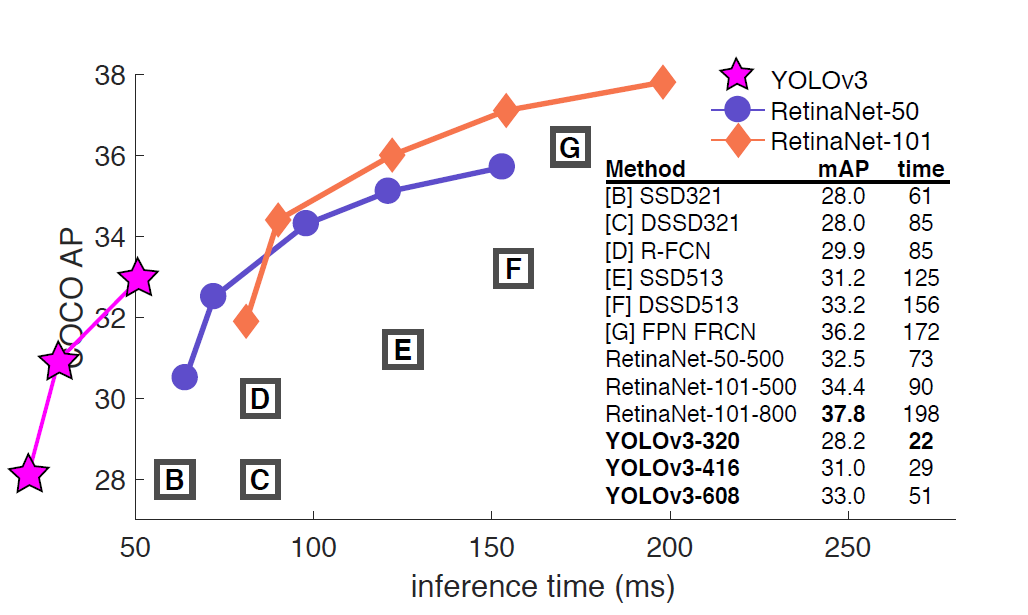
一般情况下，Two-Stage算法在准确度上有优势，而One-Stage算法在速度上有优势。其中，YOLO v3算法的训练可通过大量样本数据信息得到适用的网络参数，故我们采用业界普遍认可的TensorFlow和YOLOv3进行图片识别和图片处理，并且很多实现安全帽识别的方法在网上均有开源的框架和代码供大家参考学习，可实现生产环境安全帽佩戴情况的实时检测，因此本项目可以得到很好的技术支持。

## 5.2 行业分析

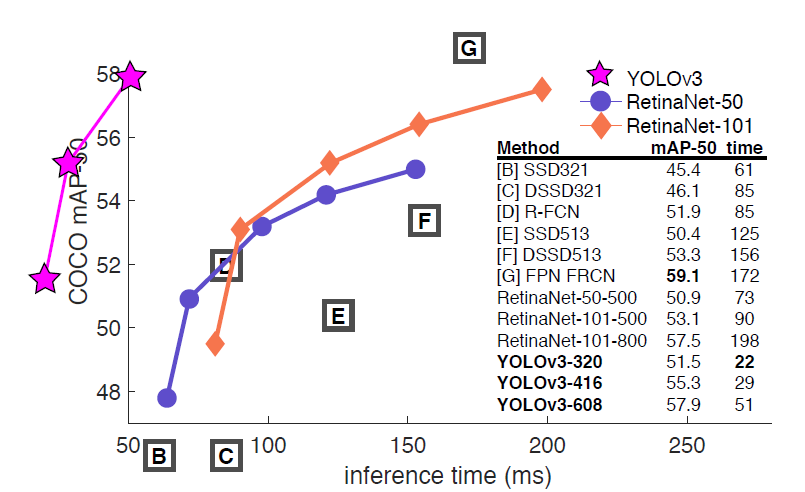
随着人工智能与各科技领域的深度融合，目标检测方法已经在机器学习、语音识别、计算机视觉等领域产生出了众多创新解决方案。基于候选区域的目标检测器，包括Fast RCNN、Faster RCNN和FPN等，以及YOLO、SSD和RetinaNet等在内的单次检测器都是目前最为优秀的方法。下表2例举了其中算法的优缺点以及适用场景，同时，对于识别算法而言，其准确率、识别速度、召回率往往难以兼顾。YOLOv3在实时检测的同时，又不会损失太多精确度，目前广泛地用于目标检测当中，有较好的应用前景。

表2 各算法优缺点比较

|  |  |  |  |
| --- | --- | --- | --- |
| 算法 | 优势 | 劣势 | 适用场景 |
| RCNN | R-CNN利用候选区域方法创建多个ROI，准确率高 | 速度很慢，内存占用量很大 | 行人检测 |
| Fast RCNN | Fast-RCNN 不会重复提取特征，因此它能显著地减少处理时间，速度比RCNN有所提高 | 存在选择性搜索，找出所有的候选框，非常耗时，速度很慢 | 行人检测 |
| Faster RCNN | 识别准确率较高，漏检率较低 | 耗时，识别速度慢 | 行人检测，[文字检测](http://www.baidu.com/link?url=bOACpk-MDnYuaSZVYDr20BYr7NVd50smLQ6FjWTL4mdhLmOZbEsrlH37ByJXJj6US2icIuRxXuJ5VvFjioSHo_) |
| R-FCN | 分析了分类与检测的位移敏感性，通过减少每个ROI所需的工作量实现加速 | 只是在一个特征尺度上进行，检测过程中会有遗漏现象 | 手势识别，人脸识别 |
| SSD | 速度快，提升小目标识别能力 | 存在一定训练难度，无法定位在低分辨率中难以检测的小目标 | 应用于大物体的识别，行人检测 |
| DSSD | 增强特征提取能力，高分辨率图片的检测精度比SSD提升明显。 | 速度降低比较明显 | 应用在小物体或稠密物体的场景中，检测具有特定关系的物体 |
| FPN | 主要解决的问题是对小物体的难以捕捉问题，目标位置准确 | 目标检测到的位置比较粗略 | 检测场景中小尺寸的物体 |
| RetinaNet | 主要解决的问题是一步做完却会导致类别不平衡问题。 | 存在极端的类别不平衡问题 | 可以检测到尺寸极小目标，航拍物体检测 |
| YOLOv3 | YOLOv3较YOLO、YOLOv2速度快，同时又能够保持较高的准确率，背景误检率低，通用性强 | 识别物体位置精准性差，召回率低 | 用于图像和视频检测，以及摄像头实时检测，行人检测 |



mAP-50



COCO-mAP

# 风险分析与对策

## 6.1 风险分析

### 6.1.1 技术风险

（1）项目预期效果风险

该种技术风险主要体现为项目设计中技术的不过关使得无法满足车间需求。如团队成员不熟悉新的技术方法、团队成员技术运用不够、团队成员所学经验不足等，这些因素使得项目不够完善，不能达到预期效果，导致技术应用失败的风险。

（2）项目技术替代风险

目前TensorFlow和YOLOv3技术已经得到了较为广泛的使用，因此很有可能在推出该方案前，现有竞争者已经攻克了该项技术。这种情形下，项目技术优势将会荡然无存。

（3）项目质量保障风险

该项目在实时监测的过程中可能存在对于运行过程不能严格把关的情况，造成出现未准确检测的情况。

### 6.1.2 市场风险

（1）潜在进入与行业现有竞争者风险

新兴技术产品的生命周期较短，更新换代快。车间安全帽佩戴实时检测系统作为市场上面出现的创新性产品，竞争激烈，市场上会不断有新的技术涌现。

（2）产品忠诚度风险

对于车间工作人员，他们有自己的工作流程，对其工作环境产生了忠诚度，一般不会轻易变化，由此对新的技术容易产生抵触情绪。

### 6.1.3 管理风险

（1）资源风险

由于本项目小组仅有5名人员，且每名人员有各自负责领域，真正实现核心技术的开发人员有限，人力成本上升和高素质人才不足。

（2）团队管理风险

一个项目成功与否，团队的配合管理尤为重要。团队成员可能由于沟通不到位和对任务理解不到位等原因，对项目实施进程造成影响。因此，队长的管理协调和队员的积极配合可以有效避免管理风险。

### 6.1.4 安全风险

（1）代码泄密风险

项目产品本身具有创新性，产品本身的核心技术保密尤为重要。产品开发安全意识薄弱，忽略智力成果的保护，如运用代码托管平台等，很可能会导致产品和新技术的泄密，研发系统被他人窃取。

（2）工具使用风险

项目研发过程中，所必须用到的管理工具、 开发工具、 测试工具等版本是否符合项目要求。

（3）系统运行风险

系统赖以运行的硬件环境和网络环境的建设进度对软件系统是否能顺利实施具有相当大的影响。

## 6.2 风险应对对策

### 6.2.1 风险应对原则

本项目坚持风险最小化、预防为主、控制为辅原则，在分析相关风险的可能性和影响程度的基础上结合风险承受度，权衡风险与收益，确定风险应对策略，并将风险管理策略贯穿整个项目研发和实施周期，强化团队成员的风险意识，遵循从风险分析、风险评估、风险预防与控制的流程，最大限度减少会给项目造成不良影响的各种风险。

### 6.2.2 技术风险对策

（1）针对产品预期效果风险

针对本产品，应该在其上市前充分进行风险评估和可行性研究，准确估量系统功能，降低技术开发风险，提高市场认可度。

（2）针对技术替代风险

产品研发过程中，公司将不断提高产品性能，丰富产品功能，并及时申请专利，提升产品的仿造门槛。在本产品研发成功之后，加大投入促使技术成果的转化，从而取得市场的主导权，降低技术成果转化风险。同时，如果有其他竞争者盗取智力成果，我们将积极运用法律手段维权。

（3）针对产品质量风险

为防范产品出现的质量风险，本团队将不断提升研发管理等各个方面的能力，以减少供应不稳定带来的风险。

### 6.2.3 市场风险对策

（1）针对潜在进入者与行业内现有竞争者风险

为了减少潜在进入者与行业内现有竞争者的风险，本团队将不断研发新产品、改革新技术，保持产品的创新性和不可替代性，使之在同类竞争产品中具有竞争优势。

（2）针对产品忠诚度风险

为降低产品忠诚度风险，本团队会对车间等市场充分调研。对车间工作人员加大宣传，使工作人员充分了解本项目，认识其他功能，并为之提供可行性报告公司。

### 6.2.4 管理风险对策

（1）针对资源风险

团队队长会合理分配每位团队成员的任务安排，确保每为成员工作量均衡，对可以投入的开发人员做到高效利用，充分发挥每位成员的优势。

（2）针对团队管理风险

推行目标管理，提高管理团队的整体素质；倡导组织创新、思想创新，适应不断变化的外部环境；加强管理监督，安排合理日程。同时，项目研发过程中多与项目团队成员交流和沟通，让每位成员明晰项目角色和责任。

### 6.2.5 安全风险对策

产品具有创新性，其核心技术内容将加以保密处理，加强在此类安全细节上的关注度。比如代码托管平台仓库设为私有等。

# 7. 结语

本文使用YOLOv3模型对训练集识别率高达92.13%，每秒识别帧数达到56.0，符合工作车间实时观测检查的标准。实际上，此模型对分辨率（1024\*768）的图像识别率达到94.22%，考虑到现在大部分车间已开始采用全高清监控摄像头，我们模型的识别率必然会更高，达到更好的效果。本项目的实现，具有一定的研究价值和经济意义，为生产车间实时监测提供了技术参考。