

**本科毕业设计（论文）**

**题目:基于计算机视觉的目标跟踪移动机器人**

**学 院 机械工程学院**

**专业班级** **创新班**

**学生姓名** **任创新**

**学生学号** 201710118118

**指导教师** **梁顺可**

**提交日期** **2021年4月9日**

**华南理工大学广州学院**

**学位论文原创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名： 日期：2021年4月9日

**学位论文版权使用授权书**

本人完全了解华南理工大学广州学院关于收集、保存、使用学位论文的规定，即：按照有关要求提交学位论文的印刷本和电子版本；华南理工大学广州学院图书馆有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务；可以采用复印、数字化或其它复制手段保存论文；在不以赢利为目的的前提下，可以公布论文的部分或全部内容。

学位论文作者签名： 日期：2021年4月9日

指导教师签名： 日期：2021年4月9日

作者联系电话： 13368686821 电子邮箱：chuangxinr@gmail.com

**摘 要**

本文详细介绍了目标检测，与目标跟踪算法在机器人系统中的应用，上位机由小型计算机实时运行程序，实现实时目标检测，并计算出目标的实际位置信息，通过串口发送到机器人 STM32 主控进行控制，实现目标的瞄准，跟随等动作。

目标检测在所检测目标的尺度变化较大时，容易检测到错误的目标或者是漏检，本文使用深度学习的方法，通过训练深度学习模型，提高了目标的检测成功率，也提高目标跟随时的稳定性。

本文所使用的最主要的传感器为单目工业相机，捕捉机器人前方的场景作为图像信息，二维的图像信息中不具有目标与传感器的距离信息参数，文中提到的坐标转换算法可以将图像中的二维笛卡尔坐标转变为三维空间坐标，从而提供机器人目标跟随时的关键距离参数以及云台相对角度。

小型计算机计算能力有限，在目标检测的算法上可以存在一定的滞后性，就是说检测到的位置可能不是目标的实时位置，再加上机器人本身的控制存在误差，使得跟随的实时性能较低，本文采用卡尔曼滤波器，可以根据目标之前的状态变化，得到目标接下来状态的最优估计，从而有效解决目标检测的不实时性。

机器人采用两轴云台设计，可以保证目标在二维平面中的跟随，而通过底盘的移动来抵消两者之间的深度差，实现对目标在设定好范围内的跟随。

**关键词**：目标检测；坐标转换；卡尔曼滤波器；目标跟随

Abstract

This paper introduces the application of target detection and target tracking algorithm in the robot system in detail. The upper computer runs the program in real time by a small computer to realize real-time target detection, and calculates the actual position information of the target, which is sent to the STM32 main control of the robot through the serial port for control, aiming, following and other actions of the target.

When the scale of the detected target changes greatly, it is easy to detect the wrong target or miss detection. This paper uses the deep learning method to improve the success rate of target detection and the stability of target tracking by training the deep learning model.

The main sensor used in this paper is monocular industrial camera, which captures the scene in front of the robot as image information. The two-dimensional image information does not have the distance information parameters between the target and the sensor. The coordinate transformation algorithm mentioned in this paper can transform the two-dimensional Cartesian coordinates in the image into three-dimensional space coordinates, so as to provide the key distance parameters of the robot target at any time And the relative angle of pan tilt.

Small computer computing capacity is limited, there can be a certain lag in the target detection algorithm, that is to say, the detected position may not be the real-time position of the target, coupled with the robot's own control error, so that the real-time performance of the following is low. This paper uses Kalman filter, according to the state change of the target before, we can get the next state of the target In order to solve the problem of non real-time target detection effectively, the optimal estimation is used.

The robot adopts two axis pan tilt design, which can ensure the target to follow in the two-dimensional plane, and offset the depth difference between the two through the movement of the site, so as to realize the target to follow in the set range.

**Key words**: Target detection; coordinate transformation; Kalman filter; target following

**目 录**

**摘 要** Ⅰ

**Abstract** Ⅱ

**第一章 绪论** 1

1.1 目标检测概述 1

1.2 课题的背景和意义 2

1.3 主要研究工作 2

1.4 章节安排 3

**第二章 基础知识介绍**  4

2.1 实时目标检测 4

2.1.1 传统目标检测 4

2.1.2 深度学习检测目标 5

2.2 卡尔曼滤波算法 5

2.2.1卡尔曼滤波简介 5

2.2.2 卡尔曼滤波算法实现运动预测 6

2.3 坐标转换 8

2.3.1 PNP算法概述 8

2.3.2 二维坐标与世界坐标的转换 8

2.4 本章小结 10

**第三章 机器人系统及调试介绍** 11

3.1 移动机器人系统 11

3.1.1 简介 11

3.1.2 机器人的硬件设备 11

3.2 调试与数据分析 12

3.2.1 调试方案及程序设计 12

3.2.2 调试数据收集与分析 14

3.3 本章小结 19

**结论**  46

**参考文献** 47

**附录** 48

**致谢** 49

1. **绪 论**

**1.1 目标检测概述**

目标检测作为计算机视觉中长期存在的基本问题，是非常具有挑战性的，近十年来相关领域的研究非常的活跃。目标检测的内容是给定一张输入图像，判断在这张图像中是否有特定类别标——如猫、狗、自行车或是人类等；如果在图像中找到了特定的目标，则还需要返回特定目标在空间中的位置。目标检测作为很多复杂的高级视觉任务的基础（如语义分割、目标跟踪、事件检测等），在计算机视觉领域有很广泛的应用，其中较为人所熟知的应用包括自动驾驶、智能视频监控、机器人视觉、增强现实等。

随着深度学习技术的发展，已经在目标检测领域中有了非常多的应用，这些技术也不断的推动着目标检测技术的发展。如今我们所使用到的刷脸支付，刷脸进站等，就是人脸识别技术的应用。深度学习的计算模型很复杂，也非常灵活，使得深度学习技术能在视觉识别、目标检测等领域中有非常广泛的应用。

卷积神经网络是深度学习中很具有代表性的一种模型，在图像分类领域中，使用卷积神经网络（CNN）的AlexNet在ILSVRC 2012挑战赛中夺魁 ，使用CNN进行图像分类逐渐成为主流，后来又被提出的R-CNN算法，Fast-RCNN算法等，取得了非常大的成功，这类算法都会先在图像中寻找到很多候选区域，然后再候选区域中再进行目标检测，这样的算法在CPU设备上运行速度较慢，不能达到实时的效果；YOLO（You Only Look Once）则是一种只需要一次就能实现目标检测的算法，YOLO算法经过多次改进，到YOLO v2、v3，速度更快，检测准确率也越来越高，开源之后被再优化的版本 YOLO v4，能够部署到移动计算设备中，并且也有很高的检测效率。

**1.2 课题的背景和意义**

人工智能是目前很热门的一个话题，随着技术的发展，人脸识别、智慧出行、无人驾驶等技术已经慢慢称为生活中不可缺少的技术。京东在上海建立的全流程无人仓库，表示机器人技术已经在工业自动化领域有成熟的应用，而其中不乏AGV机器人的身影。

在校期间，我参加了学校的机器人队，随队参加了多次机器人比赛，也有幸接触到了这些新鲜的知识，我对此也非常感兴趣，也在课余时间对相关的知识做了很多的了解，所以选择了之前有工作基础的课题。

图像处理可以应用在很多领域，智能监控中的运动目标检测、医学影像中的目标分类检测、自动驾驶中的路况检测等，都对环境要求较高，也要求所设计的算法有较高的鲁棒性，这个方向的研究内容还有很多，希望能通过这个毕业设计去学习一些相关的内容。

本课题所涉及到的目标检测与目标跟踪技术，都在自动驾驶技术中有所应用，例如交通信号灯的检测，交通标志的检测等，都用到了目标检测技术，本课题涉及到较简单的单目标检测任务，算是这类任务中的基础内容。而依据图像信息对实际的机器人进行控制，其中也涉及到诸如坐标转换、位置预测等知识，也是之后学习或是工作的基础内容。

**1.3 主要研究工作**

**1.4 章节安排**

1. **基础知识介绍**

**2.1 实时目标检测**

目标检测是计算机视觉的一个热门方向，广泛应用于机器人行业、工业自动化产线、航空航天等领域，使用计算机处理任务，减少了人力成本的消耗，提高了处理任务的效率，有较好的应用前景，也有很重要的现实意义。

**2.1.1 传统目标检测**

目标检测问题就是从给定图像中找出指定的物体，又可以分为物体定位和分类两个子问题，也就是要同时确定物体的位置和物体的类别。

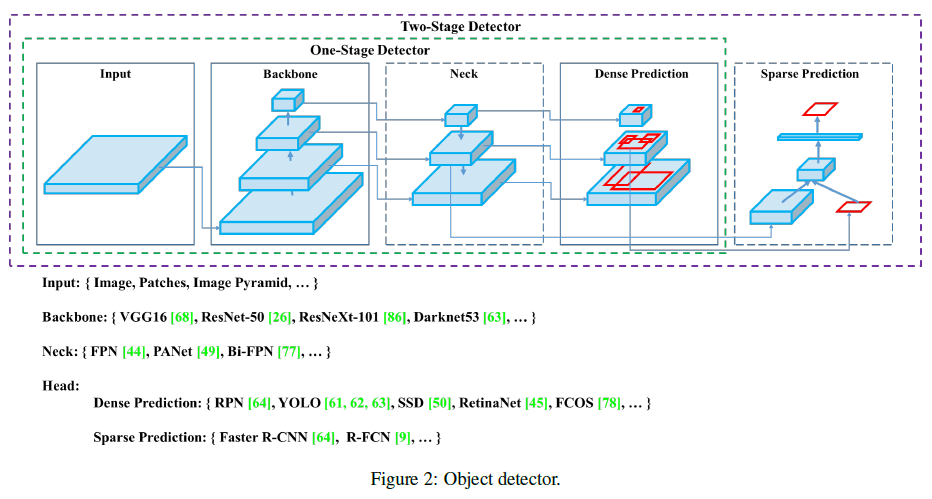
传统的目标检测方法一般分为三个阶段：

1. 首先在给定的图像上选择一些候选区域，目标在图像中可能出现在任何位置，并且无法确定目标的尺寸，所以常用方法是设置很多不同尺寸大小的区域，以滑动的方式在图像中寻找以确定物体的位置，这样的方法需要花费大量的时间，导致检测的效率低下；
2. 然后对这些区域进行特征提取，由于环境的不同，图像可能呈现出不同的颜色、亮度等，所以想要找到物体的一个鲁棒的特征是很困难的，但是物体的特征是直接影响分类的准确性的，这也是目标检测中较为困难的问题。一些常用的特征提取方法有——尺度不变特征变换（SIFT），方向梯度直方图（HOG）等；
3. 最后使用训练好的分类器进行分类，在第二步中提取出的特征将被用来进行分类，常用的分类器有SVM支持向量机等，本文不涉及，故不做过多介绍。

**2.1.2 深度学习检测目标**

随着深度学习技术的发展，在目标检测领域中也出现了很多使用了深度学习的算法，目标检测中的两个关键任务是目标的定位与分类。目前基于深度学习的目标检测算法中，大致可以分为两类：1.two-Stage目标检测算法，这类算法将目标检测问题分为两个步骤，第一步先产生候选区域，其中会包含目标大概的位置信息，然后第二步会对先前产生的候选区域进行分类，再根据分类结果对位置信息进行调整，典型的代表算法有 R-CNN、Fast R-CNN、Faster R-CNN等；2.one-Stage目标检测算法，这类算法不需要产生候选区域，直接一个阶段就产生物体分类的概率和位置信息，典型的算法有 YOLO、YOLO v2/v3等。Two-Stage算法在分类的准确度上更有优势，而one-Stage算法在速度上更快。不过随着研究的发展，两类算法都有所改进，都能在准确度以及速度上取得很好的效果。

本文主要参考了 YOLO v4算法，所以简单介绍一下 YOLO v4算法的流程，YOLO v4算法的流程如下图所示



目标图像作为 Input 输入；Backbone可以理解为图像的特征提取部分，这部分会提取一些例如颜色、纹理、边缘等特征，在网络结构上可以借鉴 VGG16，ResNet-50，ResNet-101，Darknet53等已经训练好的网络；提取特征之后，需要对提取到的图像特征进行特征增强，将前面提取得到的特征变得更加有辨识度，常见的方法有几何增强和色彩增强，几何增强包括一些拉伸、旋转操作，或者是一些随机翻转操作、随机裁剪操作等；色彩增强包括对比度增强、亮度增强以及 HSV 空间增强等；这部分的典型算法有 FPN，PANet，Bi-FPN等；然后就是整个算法最关键的部分，通过反卷积层来反卷积输出结果，接下来通过激活函数，则可对目标做出分类，YOLO v4 算法在损失函数上也有所创新，目标检测的损失函数一般由分类损失函数和回归损失函数两部分构成，而 YOLO v4 算法所使用的 CIOU\_Loss 损失函数完整的考虑了目标框回归损失函数应该考虑的三个几何因素——重叠面积、中心点距离、长宽比，使得预测框回归的速度和精度变得更快和更高。

**2.2 卡尔曼滤波算法**

**2.2.1卡尔曼滤波简介**

卡尔曼滤波适用于估计一个动态系统的最优状态，即使观测到的系统状态含有噪声，观测值不准确，卡尔曼滤波也能完成对状态真实值的最优估计。一个系统的状态有很多，例如一个水库水位的高低，某个热水壶中热水的温度等。

卡尔曼滤波认为状态空间（上一个时刻的状态到下一个时刻的状态）和发射矩阵（观测变量与状态变量）都是线性的，并且所含的噪声满足零均值的高斯分布，卡尔曼滤波是给出系统运行一段时间后的最小方差估计的算法，其中包括根据系统预测下一个时刻的状态，以及根据输入的观测值进行状态的修正两个步骤，利用增益矩阵对系统的预测值和观测值进行加权平均，减少噪声对观测值的影响，提高观测的精度。卡尔曼滤波对于干扰噪声的类型并没有假设要求，是最优的线性估计模型。

**2.2.2 卡尔曼滤波算法实现运动预测**

我们假设有一辆小车在一条笔直的公路上保持直线行驶，小车的行驶方向不会发生变化，不确定的是小车的位置和速度，卡尔曼滤波的前提条件是小车的速度和位置量在定义域内具有正态的高斯分布，即每个变量都具有一个平均值μ和方差 。速度与位移之间是有相关性的，速度越大，位移就越大，可以引入协方差矩阵来描述变量之间的相关程度。在一个系统中，会存在过程噪声和测量噪声，可以假设它们都服从正态分布，分布与两种噪声分别的协方差矩阵相关，相关的公式推导在网上以及有很多，这里做赘述，简单描述一下算法执行的步骤，首先计算出预测值的协方差矩阵、误差与真实值之间的误差协方差矩阵，通过前面的结果继续计算出卡尔曼增益的值，通过卡尔曼增益计算出估计值，再次计算估计值和真实值之间的误差协方差矩阵，将这个矩阵作为下一次递推的输入值，循环的递推，就是卡尔曼滤波算法的计算过程，最终计算出的估计值，与实际的测量值，根据一定的计算，得到最后的最优估计值。

在 OpenCV 中定义好了卡尔曼滤波的递推过程，我们需要设定好几个协方差矩阵的输入量，再将所需要进行最优估计的变量作为参数传入，输出的参数再作为下一次输入参数，之后根据实际情况调整协方差矩阵的值，以达到更加平滑的估计效果，再对一些变动较大的数据进行处理，就能得到很好的运动预测效果。

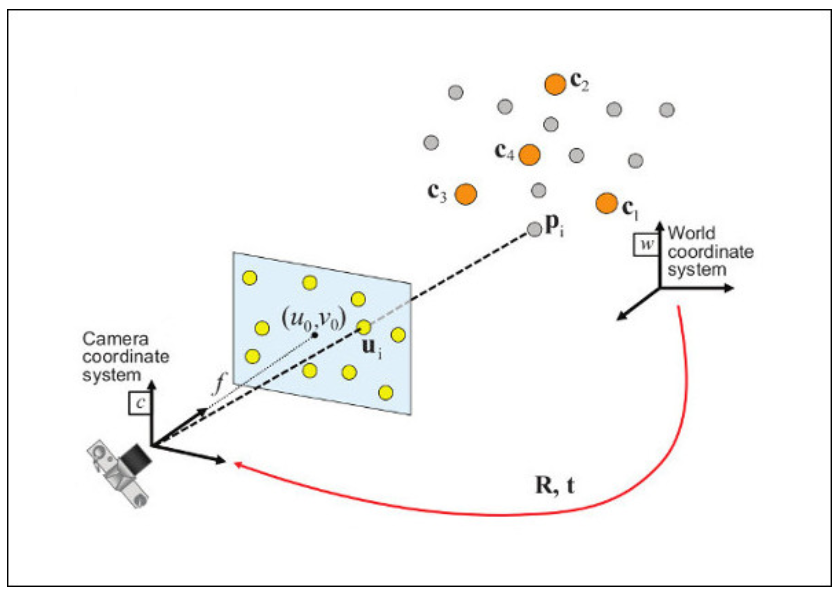
**2.3 坐标转换**

**2.3.1 PNP算法简介**

PNP 算法用于从 3D-2D 点对应关系中查找对象姿态，当知道 n 个三维空间点的坐标以及这些点在二维平面上的投影位置时，通过计算特征点与成像中的投影点，计算出其中的投影关系，从而获得相机或是物体的位姿。

通过PNP算法，可以实现单目相机的测距，这时需要用到相机内参，就是相机本身特有的一些参数，通过相机标定可以将这些参数计算出来，再应用到 PNP 算法中。小孔成像的模型可以看作是最简单的一种相机模型，模型中成像平面与小孔之间的距离可以理解为相机的焦距 f ，而成像平面就是相机的感光元件，小孔则是相机的透镜，在相机实际制作过程中，肯定会存在误差，也即是说透镜的中心与感光元件的中心并不在垂直于成像平面的直线上，所以需要引入两个新的参数 Cx 和 Cy 来表示它们之间的偏移，成像装置由多的单元组成，每个单元的尺寸都可以分为 Sx 和 Sy，从而焦距也可以分成透镜的物理焦距长度于每个的单元尺寸的分别乘积 fx 和 fy，现在我们引入了相机内参的四个参数 【Cx，Cy，fx，fy】，在实际的计算中，这四个参数将会被变换称为一个相机的内参矩阵，在实际计算时使用。而相机的外参来源于相机自身的畸变，畸变可以分成径向畸变和切向畸变，径向畸变是由透镜的形状造成的；切向畸变是由相机的安装过程造成的。径向畸变中常引入 2-3 个参数 K1，K2（K3）；切向畸变常引入 P1，P2。这里引入的几个个参数在加上相机内参，一共至少有八个参数，通过相机的标定，可以消除相机的畸变。

**2.3.2 二维坐标与世界坐标的转换**

在相机的成像中，我们常用二维的平面坐标系来表示每个像素点的位置，但是在实际中物体的位置需要用三维的世界坐标系来表示，如何把二维坐标转换为世界坐标呢，先看看下面这张图。

这样就将相机本身的相机坐标系、世界坐标系以及成像平面的二维坐标系联系了起来，通过三角形的相似性，加上我们能知道的参数，如相机的焦距，物体成像的二维坐标，以及相机的参数矩阵，就可以根据 PNP 算法计算出物体与相机的相对位姿，其中就包括相机到物体在世界坐标系中的距离。

在图像处理中，选择使用 OpenCV 作为图像处理的开源库，OpenCV 中内置了求解 PNP 的方法实现，我们只需要提前标定好相机的畸变参数矩阵，以及获取到物体的4组二维像素坐标，在将这些坐标作为输入，OpenCV 中内置的 solve\_PNP 算法就能够求解出物体的位姿信息，其中就包括我们最需要的距离信息

1. **机器人系统及调试**

**3.1 移动机器人系统**

**3.1.1 简介**

本文所设计的算法与控制，都基于华广机器人野狼队自主设计的全向移动机器人实现，该机器人主要用于参加机器人比赛，随着不断的迭代优化，以及能实现很多任务的需求，选择此机器人来作为程序设计的基础，一是有现成资源，二是熟悉开发和调试的流程，对于完成毕业设计非常有帮助。

**3.1.2 机器人的硬件设备**

本文主要研究视觉算法的设计，对机器人的机械结构以及控制元件只作简单介绍，机器人地盘采用四个麦克纳姆轮，可以通过控制四个无刷电机，实现四个轮子的速度分解，从而能让机器人实现全向移动，有利于机器人系统不需要做出很多方向上的调整，就能移动到需要的位置，对于需要稳定环境的视觉算法有很好的适用性。该机器人还拥有两轴的云台设计，水平方向上（通常称为 yaw 轴）能够实现 360°的运动，而垂直方向上（通常称为 pitch 轴）能实现俯仰共三十度的运动范围。云台同样采用无刷电机作为动力，能够实现快速的运动响应和稳定的伺服效果。云台上方安装了可以拍摄图像的工业相机，用来捕捉机器人前方的图像，在将图像传输到运算设备上进行图像算法的处理。使用工业相机的目的是因为工业相机具有很好的稳定性，长时间工作的情况下也能保持稳定的运行，并且工业相机拍摄的帧率较高，更适合用于实时性要求很高的程序上，机器人尺寸有限，所搭载的运算平台为一个迷你电脑主机，使用 Intel i5 的 CPU，能够完成大部分所需要的计算工作，工业相机通过 USB 线与运算平台连接，运算平台通过串口与机器人的 STM32 主控板连接，用于传输串口数据，实现视觉算法与控制算法的通讯，运算平台实时运行 Linux 系统，编译好的程序直接在系统上运行。

**3.2 调试与数据分析**

**3.2.1 调试方案及程序设计**

**3.2.2 调试数据收集与分析**

**3.3 本章小结**

*（1）正文层次*

*正文所有章节按“第一章、第二章、第三章……（换章时必须换页）；1.1、1.2、1.3……；1.2.1、1.2.2、1.2.3……”编排。各层次题序及标题不得置于页面的最后一行（孤行）。*

*（2）正文字号、字体、行距*

*第一级标题用三号，宋体，加粗，左右居中，上下空一行；*

*第二级标题用小三号，宋体，加粗，靠左，上下空一行；*

*第三级标题用四号，宋体，加粗，靠左，不空行；*

*正文用小四号，宋体，行距为多倍行距1.2。段首行缩进2个汉字。*

*（3）插图和插表*

*插图（表）的标题和图（表）内文字：五号，宋体。*

*插图（表）必须有图（表）题（由号和名组成）。图（表）号按章排序，如第一章第一图的号为“图1-1”。图题置于图下方，表题置于表上方。*

*插图与其图题不得拆开排写于两页。图中若有分图时，分图号用a)、b)等置于分图之下。有数字标注的坐标图，必须注明坐标单位。*

*表格不加左、右列线；表内数字空缺的格内加“——”字线；如某个表格需要换页接排，在随后的各页上应重复表的编排，编号后跟标题和（续）。如：表2-1 加入激素后的实验结果比较（续）。*

*另起页：论文主体范例*

**结 论** *三号，宋体，加粗，左右居中，上下空一行*

*正文：小四号，宋体，行距为多倍行距1.2；段首行空两个汉字。*

*另起页：参考文献范例*

**参考文献** *参考文献标题为三号，宋体，加粗，左右居中，上下空一行*

*正文为五号，宋体，行距为多倍行距1.2*

（1）学术期刊文献

［序号］作者．文献题名[J]．刊名,出版年份,卷号(期号)：起-止页码

（2）学术著作

［序号］作者．书名[M]．版次(首次免注).翻译者.出版地:出版社, 出版年: 起-止页码

（3）有ISBN号的论文集

［序号］作者．题名[A].主编．论文集名[C]．出版地：出版社,出版年：起-止页码

（4）学位论文

［序号］作者．题名[D]．保存地：保存单位,年份

（5）专利文献

［序号］专利所有者．专利题名[P]．专利国别：专利号,发布日期

（6）技术标准

[序号］标准代号,标准名称[S]．出版地：出版者,出版年

（7）报纸文章

[序号］作者．题名[N]．报纸名,出版日期(版次)

（8）报告

［序号］作者．文献题名[R]．报告地：报告会主办单位,年份

（9）电子文献

［序号］作者．电子文献题名[文献类型/载体类型]．文献网址或出处,发表或更新日期/引用日期(任选)

*另起页：致谢范例*

**致 谢** *致谢为三号，宋体，加粗，左右居中，上下空一行*

*正文为五号，宋体，行距为多倍行距1.2*