|  |  |
| --- | --- |
|  | **第十五届全国大学生软件创新大赛** |
| **文档编号：SWC2022-T20220036-抓住杰瑞** |

****

**基于多设备融合的智能补光系统+**

DeepLightening

**技术研究报告**

**Version: 1.5**

卡通人物

中度可信度描述已自动生成

**抓住杰瑞**

**2022.4.21**

**All Rights Reserved**

目录

[1 问题聚焦 1](#_Toc101476961)

[1.1 问题描述 1](#_Toc101476962)

[1.2 问题抽象 1](#_Toc101476963)

[1.3 问题评估 2](#_Toc101476964)

[1.4 问题分解 3](#_Toc101476965)

[2 相关工作 4](#_Toc101476966)

[3 技术方案 5](#_Toc101476967)

[3.1 技术方向 5](#_Toc101476968)

[3.2 技术选择 5](#_Toc101476969)

[3.3 结果期望 6](#_Toc101476970)

[4 技术实践 7](#_Toc101476971)

[4.1 使用的开发框架及依赖的库 7](#_Toc101476972)

[4.2 技术实践过程 7](#_Toc101476973)

[5 结果验证 19](#_Toc101476974)

文档修订历史

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **修订原因** | **版本号** | **作者** | **修订日期** | **备注** |
| **1** | **初版** | **1.1** | **D** | **2021.11.10** |  |
| **2** | **补充** | **1.2** | **D** | **2021.12.29** |  |
| **3** | **补充** | **1.3** | **A** | **2021.12.30** |  |
| **4** | **调整格式** | **1.4** | **A** | **2021.12.31** |  |
| **5** | **复赛补充** | **1.5** | **A** | **2022.4.20** |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

# 问题聚焦

## 问题描述

根据中国网络视听节目服务协会数据，2020年我国短视频行业市场规模达到467亿元，随着5G网络在下沉市场的降价提速，短视频和直播业务将继续迅猛发展。在短视频拍摄和直播创作中，如何运用好灯光是创作者必须掌握的一项基本技能。而灯光的运用是一门艺术又有科学性，布光讲求明暗、层次、色彩的对比，它通过从不同角度投影，不同方向构画，产生了不同的气氛。在短视频和直播拍摄中灯光所制造的氛围决定着片子的基调。拍摄时，需要调动灯光特有的能动作用，借助灯光颜色的冷暖、明暗、位置的变化、流动，结合创意成功地运用光线来完成视觉和情感的相互转换，带给观众强烈的视觉震撼和全新的艺术享受，使得短视频和直播中主题形象深入人心。

但是，短视频与直播创业团队通常规模小且结构简单，一个小型的视频创作团队或直播团队难以拥有专业的灯光师和专业的灯光布景。因此，在短视频创作和直播过程中，由于缺少了智能灯光调节，创作者们无法根据拍摄场景、拍摄时间、拍摄主题、拍摄现场的环境光等要素，巧妙运用明暗构图凸显画面重点和主题，从而造成视频主题不突出、氛围单一、效果不佳等问题。

本产品旨在通过云端协同的人工智能图像处理技术实现智能灯光调节，协助短视频与直播行业创作者完成不同主题、不同需求的内容创作。

## 问题抽象

**产品名称：**

DeepLighting你的线上灯光师

该产品需要解决的问题如下：

1. 智能灯具控制。通过接入市面上常见的智能灯具，如小米米家智能灯系列，实现对灯具的智能控制。
2. 拍摄主题选择。用户根据创作需求选择不同的灯光主题，如美食、美妆、黑暗森林、甜蜜暴击等。
3. 云端协同的智能灯光调节。通过调用用户设备摄像头获取视频，在云端通过人工智能图像处理，寻找最好的光影构图方案，以凸显画面重点和主题，并将灯具位置与灯光强度调节数据传回用户终端。
4. 多灯具协同打光。通过终端连接多个智能灯具，协同调节灯具位置与灯光强度，实现用户所需主题氛围。
5. 多机位视频处理。通过多设备多机位的视频拍摄与处理，提高灯光调节的适应力与氛围感。

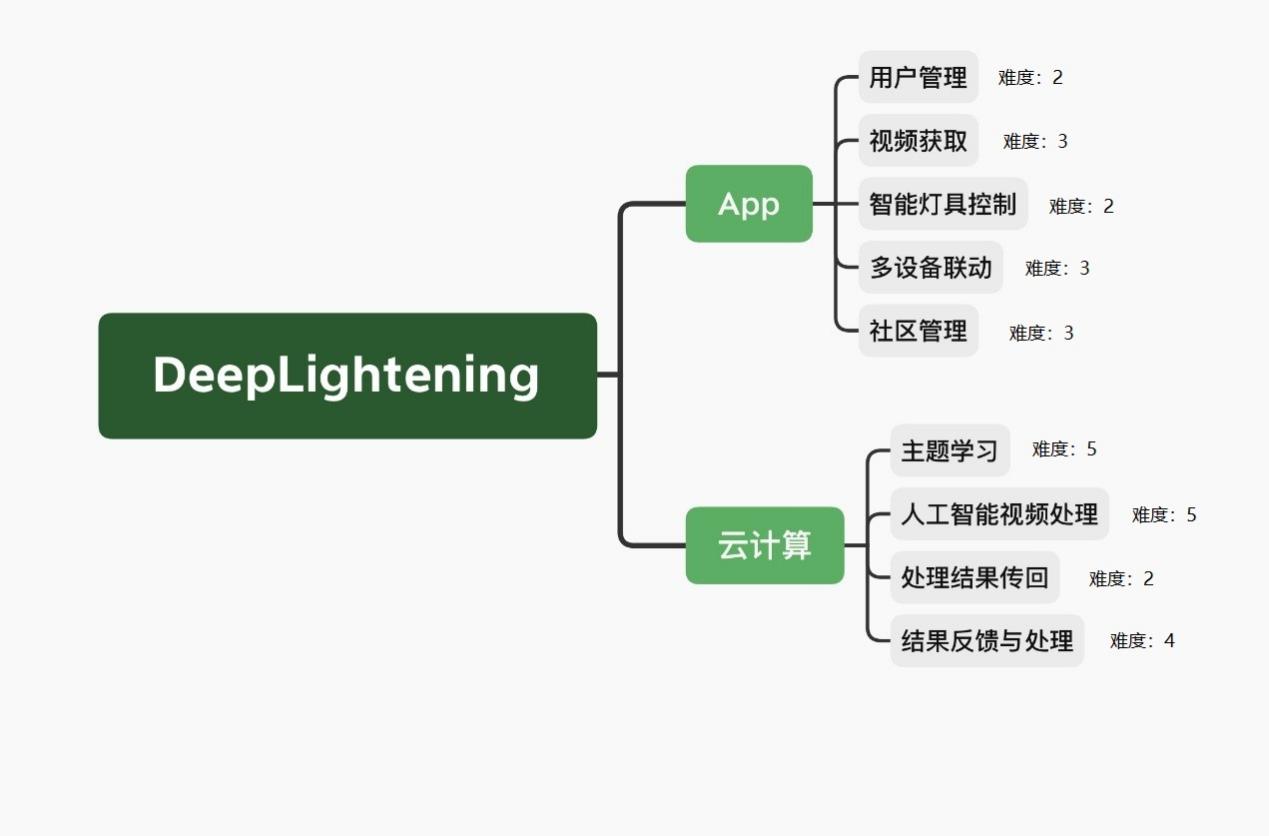


## 问题评估

随着短视频与直播行业的迅猛发展，精品化的内容创作也越来越也依赖于智能化、专业化的场景设备，智能灯具行业也随之迎来大幅增长。目前，智能灯具产品丰富，适应场景众多，在视频拍摄中使用的灯具包含了桌面灯、补光灯、摄影灯、氛围灯等，如何结合不同的拍摄主题，对多个灯具进行协同控制打光，成为多数创作者需要首要解决的难题。

基于当前智能灯具的普及与人工智能图像识别技术的发展，我们可以通过云端协同的人工智能技术进行人物识别和环境分析，基于拍摄主题与场景需求，对多主题多灯具的拍摄场景进行智能灯光调节，以满足短视频创作者、直播用户对各类场景下灯光氛围的需求，给每一个用户定制独特的灯光。

## 问题分解



# 相关工作

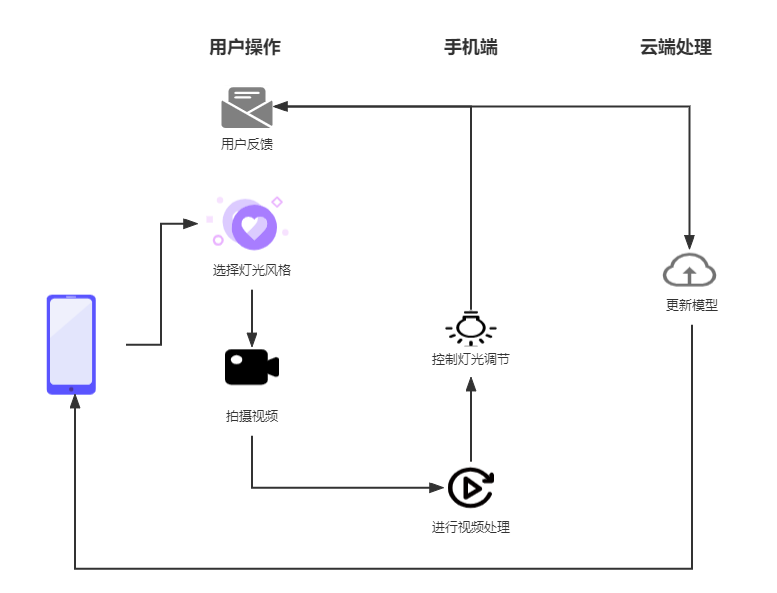


图1. 场景图

DeepLightening 基于人工智能图像识别的智能打光系统，利用人工智能控制身边的智能灯光设备，成为你的独家灯光师。

本产品旨在解决短视频、直播场景下对智能灯的控制与智能调节。

1. 用户选择灯光主题，例如美妆、科幻、带货等。
2. 拍摄视频或图片，系统将构建场景中拍摄设备、灯光设备、人物、拍摄物品的位置空间模型。
3. 计算最优的打光方案。
4. 通过局域网或蓝牙自动调节智能灯光。
5. 用户反馈打光效果。
6. 将当前模型、打光场景、用户反馈上传云端，学习更新打光模型。

图示, 示意图

描述已自动生成

图2 项目系统架构

# 技术方案

## 技术方向

深度学习、物联网、云端协同，多机位，视频压缩，分布式，灯光位置建模

## 技术选择

安卓开发API：API31，用以调用手机硬件

应用模式：分布式应用，可以提高容错率和性能

数据库：MySQL，用以存储数据

视觉库：OpenCV4.5.4，进行物体识别和光照计算

网络传输协议：HTTPS，能安全的传输数据，防止用户数据泄露

脚本：Python，快速开发相关功能

云端：阿里云服务，进行云端计算，减少用户端计算量，Nginx服务器反向代理，实现分布式部署。

建模方式：3D建模，便于计算物体空间关系和光照

开发平台：android studio，用以开发android应用

算法库：tensorflow

互联协议：wifi/bluetooth协议

后端：springboot框架

## 结果期望

1. 使用场景全面：APP在常见手机上均可正常运行
2. 覆盖wifi蓝牙多种连接方式：数据传输、通讯正常
3. 图像识别，光照计算、调整：人像、灯具识别正常，光照计算于调整符合实际情况
4. 人工智能打光：人工智能学习打光正常
5. 基于多主题风格的灯光调节实现：灯光调节效果符合预期风格
6. 从开始到智能灯响应，在网速条件为2MB/s或以上条件下不能超过30s
7. 多机位拍摄视频

其他待补充

# 技术实践

## 使用的开发框架及依赖的库

硬件方面智能设备经历了多年发展，已经产生了大量的智能终端，例如智能手机，电脑，平板电脑，智慧大屏等。同时智能灯具种类丰富，可以实现多场景，多设备互联融合。

人工智能经过数十年的发展，在软硬件方面都有了长足的发展。在准确率，速度等方面都能够满足本产品的需求。

**智能灯光控制：**1.树莓派3b+ Homeassistant

2. 接入小米关网

**图像处理方面**：本产品将会采用opencv技术+tensorflow算法库，实现智能学习，调节灯光

**安卓APP方面**：Android API31 + sqlite数据库+springboot框架

**云端处理**：借助阿里云服务器，并利用Nginx反向代理，实现分布式

**位置处理：**采用3D建模技术，获取位置坐标

**摄像头接入：**调用摄像头处理API

**即时通信：**即时通信模块借助环信开源SDK独立开发

**灯光通信协议：**采用bluetooth/wifi通信协议

**手机设备连接：**bluetooth/wifi协议

## 技术实践过程

硬件准备：

1. 工作电脑

2.树莓派3b

3. 一张16G以上tf卡、tf卡套或USB口tf卡读写卡器

软件准备：

1. 烧录软件etcher(https://etcher.io/)

2.文本编辑软件notepad++(https://notepad-plus-plus.org/downloads/)

3.Android Studio

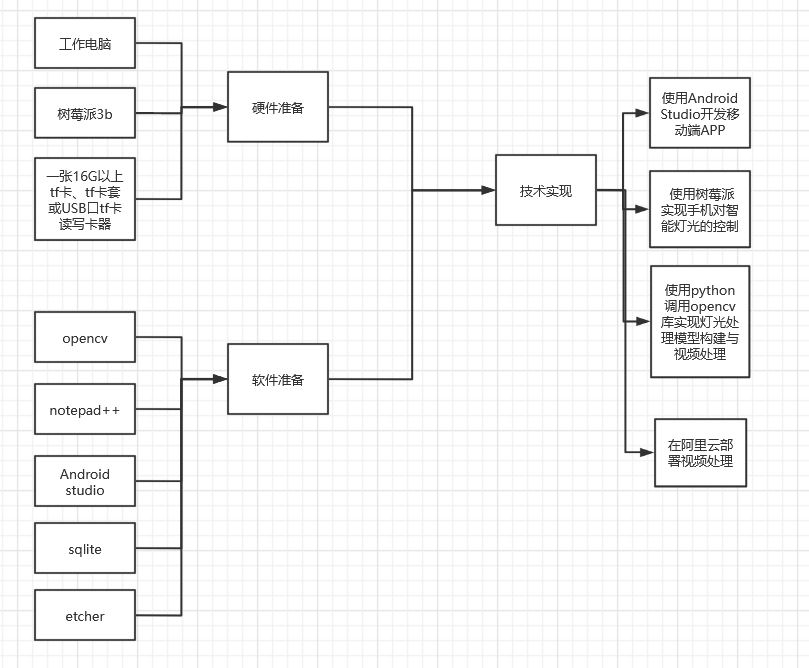
4.数据库sqlite

5.视觉库opencv

技术实现：

1. 使用Android Studio开发移动端APP
2. 使用树莓派实现手机对智能灯光的控制

3.使用python调用opencv库实现灯光处理模型构建与视频处理

4.在阿里云部署视频处理

**后端视频处理模块：**

总流程图

图示

描述已自动生成

主机位流程图

图示

描述已自动生成

次机位流程图

图示

描述已自动生成

相关过程的逻辑说明如下：

**第一部分 面部测距**

算法能够提取面部的特征，计算出面部到相机的距离。不同的人脸尽管看起来千差万别，但客观上存在一些极为相近的特征，我们可以借此做出较为精确的距离判断。这个特征就是两个瞳孔之间的距离，即瞳距。一般而言，男性的平均瞳距为64毫米，而女性的平均瞳距为62毫米，我们将63毫米作为标准。在客观上这样会存在一定的误差，不过这对于补光结果的影响是可以忽略不计的。

图片包含 图示

描述已自动生成

如图，w表示物体在图像上的宽度，W表示物体实际的宽度，在此处表示63毫米的瞳距；f为相机的焦距，d表示物体到相机镜片的距离，则容易推出d = (W \* f) / w。

但是要注意，d = (W \* f) / w只适用于人脸正对着摄像机的情况。如果观察者做出转头动作，数据会出现明显的错误。本项目在上述基础上，融入了head pose estimation技术，通过六个关键点对人脸的朝向进行了智能分析，根据面部的方向向量对瞳距进行修正，进而实现单目测距。如此，只要拍摄设备在人的前方，便可以自由进行单目测距，进而实现位置建模。

**第二部分 智能灯测距**

该部分算法同面部测距基本相似，软件通过视频或图像获取到圆柱形支架的尺寸，便同样可以得到智能灯到相机的距离。

**第三部分 位置建模**

图示

描述已自动生成

根据在图像上灯具到人脸的水平距离与竖直距离，也可以计算出灯与人脸在空间上的关系。为了简化过程，我们假设人脸在相机的正前方，以相机右侧为x轴，相机前方为y轴，相机上方为z轴，则人的坐标为（0,distance,0）,灯1的坐标为（x1,,z1）,灯2的坐标为（x2,,z2）。

则灯1到人脸的距离

d1 =

=

同理：

灯2到人脸的距离d2 =

进一步拆解该式子：

W为实际瞳距，w为图像上的瞳距。

dx1为图像上灯1到人脸的水平距离，dz1为图像上灯1到人脸的竖直距离，

w1为灯具的某一部分在图像上的宽度,W1为灯实际的宽度。

则可以借助智能灯的大小来判断出实际的距离，具体计算方式为：

x1 = (dx1 \* W1) / w1

z1 = (dz1 \* W1) / w1

而灯1到相机的距离distance1 = (W1 \* f) / w1

人脸到相机的距离distance=(W \* f) / w

代入得到d1 = ()

同理可得d2= ()

至此，两个智能灯到人脸的距离d1,d2就算出来了。

关于方向向量：

以灯1为例，灯1到人脸的方向向量为（-x1,distance, -z1）即（），代入具体数值即可判断出灯相对于面部的具体方位。

位置建模还借鉴采用百度视觉技术部开源的时空建模技术。深度学习在静态图像理解上取得了巨大成功，但是针对视频时空建模中最有效的网络架构是什么还尚不清楚，因此我们将新探索的用于视频中局部和全局时空建模的时空网络（StNet）架构与现有的CNN+RNN模型或是基于纯3D卷积的方法进行比对分析，来寻求更有效的网络架构。

空间建模方法对比：

**现有方法分析**

由于深度学习在图片识别中的卓越表现，该技术也被应用到了解决视频分类的场景当中。这其中就有两个主要的研究方向，一个是应用CNN+RNN框架结构来对视频序列建模，还有一个是单纯的利用卷积网络结构来识别视频当中的行为。但是在动作识别准确性方面，目前的行动识别方法仍然远远落后于人类表现。现有方法存在如下待改进之处。

CNN**+**RNN**模型**

对于CNN+RNN的方法，CNN前馈网络部分用来空间建模（spatial modeling），LSTM或者GRU用来时域建模（temporal modeling），由于该模型自身的循环结构，这导致了端到端的优化困难。单独训练的CNN和RNN部分对于联合的时空特征表示学习（representation learning）不是最佳的。

**纯卷积网络结构**

2D卷积网络结构在抽取外观特征（appearance features）的时候，只利用了局部的空间信息而忽略了局部的时域信息；此外，对于时域动态，2D卷积网络仅融合了几个局部片段的分类得分并计算平均值，这种取平均的方法在捕捉时空信息方面的性能有待提高。3D卷积网络结构可以同时在空间和时间上建模进而得到令人满意的识别任务结果。众所周知，浅层的神经网络与深层神经网络相比，浅层网络在大数据集中，表现出较差的表示学习能力。当进行大规模数据集中的人类行为识别任务时，一方面浅层的3D卷积网络得到的视频特征的可辨别性相对深层网络较弱，另一方面，深层的3D卷积网络会导致过大的模型以及在训练中和推理阶段中过高的计算成本。

**图示

描述已自动生成**Stnet**模型**

基于ResNet骨架构建的StNet。StNet的输入是T ×3N ×H ×W张量。通过2D卷积对局部时空模型进行模型。在Res3和Res4块之后插入时序卷积模块进行全局时空特征建模。最后，用时序Xception模块进一步建模时序动态信息。3D卷积的设置是（# Output Channel, (temporal kernel size, height kernel size, width kernel size), # groups) -(Ci, (3,1,1), 1)。

**超图像（**Super**-**Image**）**

StNet的输入为均匀采样的T个局部连续N帧的视频帧。局部的连续N帧组合成一个“超图”，这使得“超图”保留原始视频各个局部的时空信息。所以网络的输入是一个尺寸为T\*3N\*H\*W的张量。

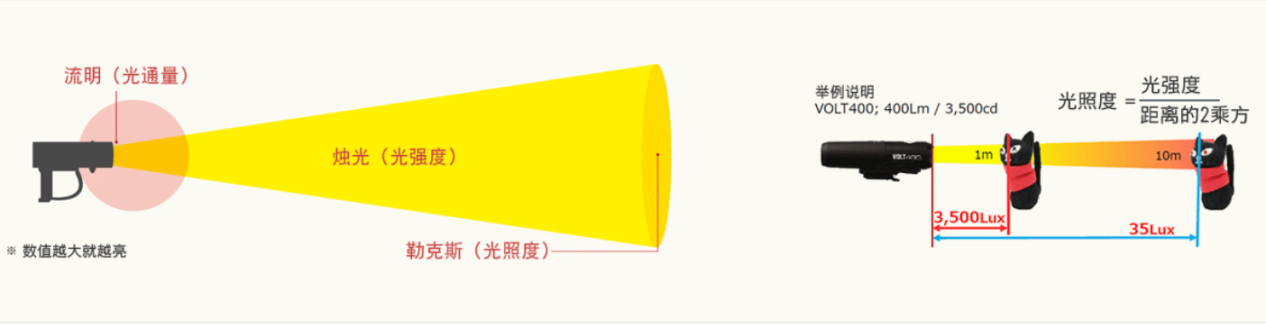
**时域建模块（**TemporalModelingBlock**）**

采用2D卷积对T个“超图”进行局部时空关系的建模，可以避免 3D 卷积网络参数量和计算量大的问题，进而生成T个局部时空特征图。通过堆叠3D卷积/2D卷积模块，对T个局部时空特征图进行全局时空信息的建模，这对理解整个视频起到至关重要的作用。具体而言，我们选择插入2个时域建模块在Res3和Res4块之后。时域建模块是为了捕捉视频序列内的长期时域动态，可以利用Conv\_3d-BN3d-RELU架构实现。将3D卷积空间维度的kernel size设置成1以节省模型的参数量与计算量。

**时域**Xception**模块（**TemporalXceptionBlock**）**

时域Xception模块是为了在特征序列之间进行有效的时域建模，并能轻松地进行端到端优化。Xception模块的设计主要基于时序1维卷积，采用了channel-wise和temporal-wise分离的策略进一步减少计算量与模型参数量。

**第四部分 光照强度的提供逻辑**



光的照度E，描述的是照到物体表面的光照强度，光的照度与距离L的平方成反比。光照到人的脸上发生漫反射，这会损失部分光的能量，这需要考虑在内。而由于光的漫反射的特殊性，我们不需要考虑漫反射后光因为传播而损失的能量，即不必考虑光从人脸到达摄像机的过程。

我们设定智能灯的光照强度为I，物体距离为L，则光照度为E = I/，我们获取到视频中人脸面部的RGB值，通过表达式0.241 \* (R^2) + 0.691 \* (G^2) + 0.068 \* (B^2) 即可得出当前面部的亮度。我们设定面部对光的反射率为R，则实际光强对人脸面部亮度影响为ΔB=，注意，这里的亮度影响人脸的位置和灯的布局和角度有关，比如灯在人的左侧，则受影响的是左侧面部。

我们将人脸划分为四个区域，以鼻子为中心，依次分为左上、右上、左下、右下。在位置建模之后，智能灯与人脸的相对关系已经得出，那么对于每一盏灯，其对人脸的四个区域的亮度影响梯度便是可知的。我们也会预先设置好正常补光下的面部亮度数据，这个亮度数据也可以通过用户的反馈进行自动地调整。基于此，我们便可以给出三种方案：第一种方案是优先满足一号智能灯的高梯度区域的照明，第二种方案是优先满足二号智能灯的高梯度区域的照明，第三种方案是两具智能灯均满足最低梯度的照明。根据ΔB=，结合第三部分的建模得出的方向向量和距离L，软件便可以计算出三种不同角度下的光强调整方案了。

**第五部分 灯光模式的设置**

根据用户选择的模式、输入的视频以及灯光的参数，首先利用选择的模式选择灯光的rgb范围，使用之前得到的灯光rgb范围和灯光强度在这一范围内进行迭代计算灯光效果，通过提前训练好的神经网络分类器对图像模式进行类型评定，在类型分数超过阈值时输出当前的rgb值，否则根据类型评分偏向进行迭代计算。

上文提到的神经网络分类器是tensorflow+cnn实现的分类器，数据部分源于网络图片，部分源于各大平台直播内容爬取，数据标注由人工标注。

对于脸部每个像素颜色在HSV颜色空间中H，即一般黄种人肤色，下文称人脸，对其RGB值加上L0并规范化。其中L0为环境光，L1，L2为用户灯，L0=（L1+L2）\*0.02，规范化指像素min(255,max(0,待规范的像素值))。记像素和灯坐标对应线段的向量为q，如果<n,q> >= range，则不处理该像素，记像素加值c=，其中d为像素到光源的距离。根据人脸和灯具的位置关系，决定人脸光照系数s。对于每个待处理的像素，其RGB+∑c\*s。最后像素RGB值规范化。对每个风格预设一个灯光的RGB，记为model。将上述图片使用cnn分类器进行分类，得到风格的置信度数组output。若max（output）对应的值为用户选择模块，返回L1，L2。否则对于每个灯L，L=L-（（非预选风格model\*置信度）+预选风格model\*置信度） \*path并规范化。若计算次数过多，返回error。否则继续迭代。

上述cnn图片分类器是使用tensorflow编写的，数据集大小约为10000，数据大部分来自美图秀秀，各大平台直播截图，其余为知名调光师所提供的调光图。

灯光模式的训练以及分析的过程如图66日程表

描述已自动生成

图66 灯光模式的训练以及分析的过程

1. **主次机位的视频/图像的用途**

主机位拍摄的内容用于人脸初始亮度分析、人脸位姿判断、灯具位置判定，并确定补光方案。

次机位拍摄的内容用于记录直播过程中的人脸亮度变化情况，灯具位置。

**限制条件**

1 .本功能模块需要识别面部与灯具，为确保成功识别，使用场景不宜过暗。

2 .拍摄时设备应尽量与人脸平行，且不应在人的背面的拍摄。

3 .需要有至少2Mb/s的带宽来确保用户的体验。

**蓝牙模块研究**

1. **硬件设计**

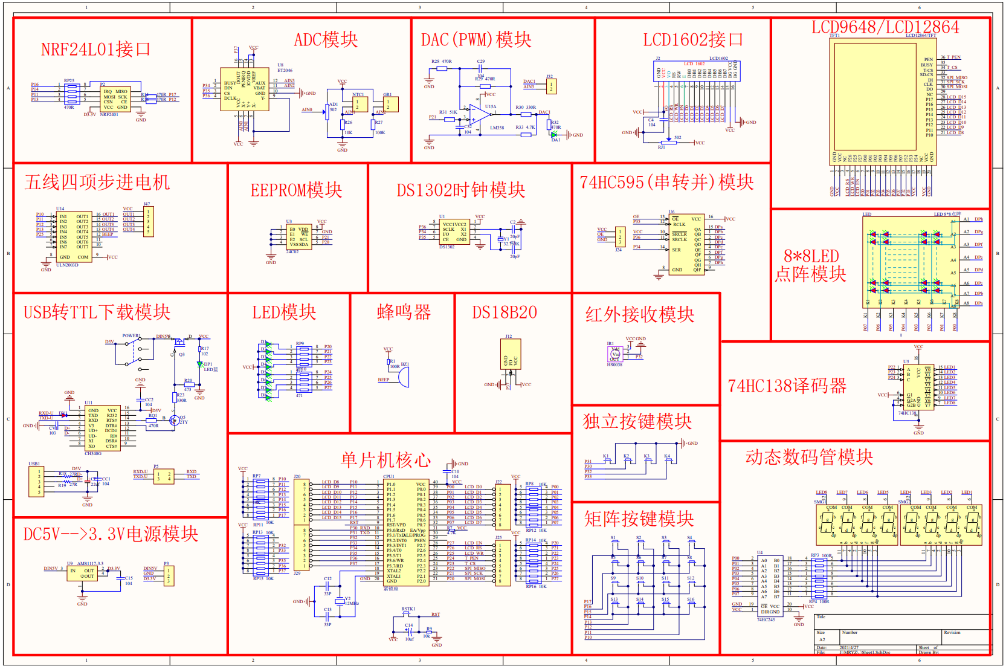


图2 普中-51单片机开发板原理图

图片包含 图示

描述已自动生成

图3 蓝牙模块

1 .供电：通过普中-51单片机开发板电源模块供电。外接3.7v电源为灯板供电。

2 .蓝牙连接：蓝牙通过RXD接[单片机](https://so.csdn.net/so/search?q=%E5%8D%95%E7%89%87%E6%9C%BA&spm=1001.2101.3001.7020" \t "_blank)P3^0，蓝牙TXD接单片机P3^1，蓝牙GND接单片机GND，蓝牙VCC接单片机3.3v。

3 .灯板连接：灯板正负极连接补充3.7v电源正负极，灯板信号线连接单片机P0^1 。

1. **备选硬件设计**
2. 芯片：采用ESP8266芯片
3. 软件：采用WS2818开源项目
4. **软件设计**
5. 通过蓝牙接收数据，数据接收完毕，信号灯闪烁。
6. 依据亮度要求以及颜色要求，相应位置灯珠以相应RGB颜色亮起。
7. 断开灯板供电，灯板熄灭。

图片包含 图示

描述已自动生成具体调节信息如下图

智能灯具调节

# 结果验证

五种色调风格，其中补光色调如下：

戴着眼镜的人

描述已自动生成 桌子上的玻璃花瓶

描述已自动生成

清新自然光(效果1) 清新自然光(效果2)

小孩手里拿着手机

中度可信度描述已自动生成 小孩戴着眼镜

描述已自动生成

温馨暖色光(效果1) 温馨暖色光(效果2)

穿蓝色衣服的人

中度可信度描述已自动生成 图片包含 人, 桌子, 室内, 食物

描述已自动生成

高级蓝色光(效果1) 高级蓝色光(效果2)

男人拿着手机

中度可信度描述已自动生成 男人伸出手

中度可信度描述已自动生成

经典红色光(效果1) 经典红色光(效果2)

人拿着手机

中度可信度描述已自动生成 穿白色衣服的人

描述已自动生成

可爱芭比粉(效果1) 可爱芭比粉(效果2)