|  |
| --- |
|  |
| “点彩成乐”-基于机器学习和画面识别的智能配乐  商业计划书 |

目录

[摘要 4](#_Toc145692346)

[一、项目概述 4](#_Toc145692347)

[1.1项目背景 4](#_Toc145692348)

[1.2项目意义 5](#_Toc145692349)

[二、项目组概况 6](#_Toc145692350)

[2.1项目组概述 6](#_Toc145692351)

[2.2发展规划 6](#_Toc145692352)

[三、产品与研发 7](#_Toc145692353)

[3.1产品简介 7](#_Toc145692354)

[3.2市场定位 7](#_Toc145692355)

[3.3产品特色优势 8](#_Toc145692356)

[3.3.1新颖性 8](#_Toc145692357)

[3.3.2竞争优势 8](#_Toc145692358)

[3.4技术研发水平 9](#_Toc145692359)

[四、市场营销 10](#_Toc145692360)

[4.1市场分析 10](#_Toc145692361)

[4.2 SWOT分析 10](#_Toc145692362)

[4.2.1优势 10](#_Toc145692363)

[4.2.2劣势 11](#_Toc145692364)

[4.2.3机会 11](#_Toc145692365)

[4.2.4威胁 11](#_Toc145692366)

[4.3营销策略 11](#_Toc145692367)

[4.4盈利模式 12](#_Toc145692368)

[五、团队介绍 12](#_Toc145692369)

[5.1导师介绍 12](#_Toc145692370)

[5.2成员介绍 12](#_Toc145692371)

# 摘要

当今社会，互联网飞速发展。人人分享的互联网时代已经到来。据统计，我们的调查对象有高达69.81%已经离不开社群分享类视频平台。而在这些人群中，又有仅仅8.33%的人没有遇到过视频配乐选取问题，有91.67%的人群在进行图片、视频配乐时遇到过各种各样的问题。我们发现为图片、视频进行人工智能配乐是一片尚未开发的蓝海。同时，不少音乐创作者面临创作时失去灵感，需要借助工具继续创作的问题。而我们的“点彩成乐”可以很好地解决这个问题。

我们开发了名为“点彩成乐”的基于机器学习和画面识别的配乐程序，可以根据图片或者视频的情绪以及其他要素，依托我们的人工智能模型生成一段图片的配乐。通过使用我们的配乐，普通用户可以为自己的视频配上专属于自己的音乐，杜绝了抖音等平台的配乐选择系统带来的同质化的抖音神曲。专业用户也可以使用我们的系统辅助自己的创作。

**关键词：**音乐生成 图像识别 深度学习 人工智能 社交网络 音乐创作

# 一、项目概述

## 1.1项目背景

近年来，随着移动互联网的飞速发展，视频分享社群网络如雨后春笋般出现，极大程度上丰富了网民的生活，不仅拓宽了网民的视野，也吸引了更多人加入短视频行业。CNNIC发布的最新数据显示，2018-2022年，我国短视频用户规模持续增长。2022年上半年，短视频的用户规模增长明显，较2021年12月增长2805万至9.62亿人，增长率达3.0%，带动网络视频的使用率增长至94.6%。随着用户规模进一步增长，短视频与新闻、电商等产业融合加速，信息发布、内容变现能力逐渐增强，市场规模进一步扩大。

然而，随着短视频平台创作者的大量涌入，内容同质化的问题也逐渐出现。据我们调查，视频配乐同质化的问题尤为严重。这种同质化问题可能会导致视频曝光量点赞量等数据不佳，使得创作者收益减少，也可能使得部分内容充实但配乐重复的视频不能获得应有的曝光。同时，根据我们的调查，在短视频创作者中，又有仅仅8.33%的人没有遇到过视频配乐选取问题，有91.67%的人群在进行图片、视频配乐时遇到过各种各样的问题。这说明视频配乐的生成是一个亟需解决的问题。

同时，随着移动互联网和相关技术的发展，发布音乐和创作音乐的门槛降低，我们也希望为新兴音乐创作者提供一个音乐辅助创作工具。

基于此，我们项目组着手开始了“点彩成乐”-基于机器学习和画面识别的智能配乐的项目开发。

## 1.2项目意义

本项目结合了图像识别和音乐生成相关的最新前沿技术，为图片、视频配乐和音乐创作提供了新的思路。同时本项目组通过自主探究，对相关技术的研究有一定贡献。

本项目研发的“点彩成乐”产品能够解决目前短视频配乐同质化的问题，在提升短视频质量的同时，能够提升视频的曝光量和创作者的知名度，助力创作者更好地创作视频内容，让创作者将精力更多地放在内容上，而不是苦恼视频配乐的选择。助力短视频创作也可以促进第三产业发展，促进县域经济的发展，促进人民就业。

同时，本产品对于音乐创作者的创作也有一定的帮助，也可以吸引更多人参与到音乐创作中。

# 二、项目组概况

## 2.1项目组概述

2022年诞生于北京， “点彩成乐”为音乐创作提供了一种新的可能性。本项目致力于使用AI深度学习的方式去解决图像/视频配乐问题。我们通过自研的AI模型将图片转换为音乐，为用户提供音乐生成服务。

目前项目组由八位各有所长的大学生组成。本项目组组员均有人工智能相关知识以及开发经验，有很强的学习能力和对前沿技术的追踪能力，并对绘画或音乐各有一定造诣。本项目组有经验丰富的程序前后端开发人员和市场调研分析人员，能开发出易用的前端和功能强大的后端，并能主动追踪市场动向，从而实现完整的商业闭环。

## 2.2发展规划

目前本项目组通过阅读大量文献和探究可行的方法，已经确定了技术路线，通过提取图片特征，将图片特征转换为文字，将文字提炼为关键词，将关键词转换为音乐。我们计划沿用继续完善现有的模型，做出完整的前后端程序，并进行小范围内测。

待内测结束后，我们计划进行广告投放和宣传，将程序上线进行公测，收集用户体验，并根据用户反馈意见继续优化我们的模型和前端程序。

# 三、产品与研发

## 3.1产品简介

我们的产品包含了完备的前后端程序以及相关预训练模型。用户在前端提交图片或者视频，并选择偏好的情感色彩，点彩成乐就能通过我们引以为傲的系统框架生成一段美妙的旋律供用户选用。在后续的开发中，我们计划继续提升我们的生成效果，并逐步开放API，为社群分享平台提供图片、视频配乐服务。

我们的愿景是，让每个人都成为作曲家。

## 3.2市场定位

我们将自己定位为以下层次

1.音乐创作工具：我们使用了先进的AI生成算法，能够使用我们经过论证的技术路线将图片转换为音乐。音乐制作人可以将自己拍摄的照片等图片上传到我们的产品，我们的产品就能为音乐创作人生成一段合理的音乐，帮助作曲家进行作曲工作，找到作曲灵感。

2.配乐生成工具：我们的AI算法可以瞄准更加广阔的市场。根据市场调研，有高达69.81%已经离不开社群分享类视频平台。而在这些人群中，又有仅仅8.33%的人没有遇到过视频配乐选取问题，有91.67%的人群在进行图片、视频配乐时遇到过各种各样的问题。我们可以为这些人群提供无版权配乐生成服务，杜绝千篇一律的“抖音神曲”，为视频和图片分享者提供令人眼前一亮的配乐。

3.技术解决方案提供商：在我们的技术经过生产环境验证后，我们可以直接接入视频分享平台以及社交平台，为他们提供配乐生成服务，提高视频分享平台以及社交平台的客户满意度以及留存度。

## 3.3产品特色优势

### 3.3.1新颖性

技术亮点：我们使用了深度学习相关领域和生成式AI的最新成果，对于音频生成大模型有一定的研究与使用。同时我们使用了图像识别在人工智能领域的最新成果，将二者结合，得到了我们的产品。

应用亮点：当人们在社交媒体上发布图片或视频时，系统可以根据其中图像的特点自动生成音乐。同时，这款产品也可以在各种新媒体中得到广泛应用。现今传统媒体大多开设了自己的网络媒体账号并发布短视频作为新闻报道。但这往往导致短视频的背景音乐趋向同质化，同时也出现了许多视频内容与音乐并不匹配的情况。本产品的应用可以有效地规避此类情况，通过深度学习的方法为各类视频提供最合适的音乐。同时，我们的产品可以为分享类短视频进行配乐，亦可以为专业人士提供创作灵感，或者提供技术解决方案。

商业亮点：本产品可以通过贴片广告、提供版权、提供API和技术支持等方式盈利。

### 3.3.2竞争优势

ChatGPT只能根据用户提供的关键词完成作词工作。只有词没有曲的作品不能称之为音乐作品，但是拥有曲调没有歌词的作品可以称作音乐作品。当询问ChatGPT能否根据图片情感作曲时，他给出了否定的回答，这说明了ChatGPT没有音乐创作能力，而我们的产品有音乐创作能力。

目前较为先进的Suno AI虽然可以创作出动听的音乐，但是该产品需要用户自己思考音乐提示词来进行音乐生成。用户不能通过直接输入图片或视频来生成自己想要的音乐，使用门槛较高。我们的产品能直接将用户提供的图片或视频转换为音乐，使用门槛低，生成质量可以满足任何图片视频的配乐要求。这使得用户在为自己的短视频或图片集配乐时，不必绞尽脑汁寻找配乐或想出生成音乐的提示词。

## 3.4技术研发水平

团队拥有自研“点彩”框架，通过三个模块实现多模态生成。

一、通过BLIP预训练模型，将用户输入的图片/视频转化为画面信息描述文本。

二、将画面信息描述文本输入大语言模型，并指定提示词进行调整，将文本转化为倾向描述音频的文本。这一模块支持用户直接输入文本。

三、通过调整后的MusicGen模型根据这段文本生成适配的音乐。

基于“点彩”框架，团队开发了“点彩成乐”系统。系统包含完善的后端与可交互的前端网站，可以为用户提供图片转音乐的服务。我们同时还积极进行B端合作方式，通过开放API、提供技术支持等方式向B端提供音乐生成服务。

# 四、市场营销

## 4.1市场分析

根据我们的市场调研，我们的调查对象有高达69.81%已经离不开社群分享类视频平台。而在这些人群中，又有仅仅8.33%的人没有遇到过视频配乐选取问题，有91.67%的人群在进行图片、视频配乐时遇到过各种各样的问题。我们发现为图片、视频进行人工智能配乐是一片尚未开发的蓝海。

同时，我们也发现一些音乐创作爱好者在创作之时遇到困惑，在作曲到中途的时候突然失去灵感，感到非常的痛苦。我们的产品也能部分地解决这个问题。

## 4.2 SWOT分析

### 4.2.1优势

我们使用了目前较为火热的人工智能技术和一些大模型相关的技术，追踪了目前的技术热点，使用新技术和新思路解决了困扰创作者已久的老问题。同时，我们有较为成熟的商业应用场景，推广难度虽有但是并不大。

项目组成员对目前前沿的人工智能技术均较为了解，并均有一定的艺术造诣，对数字音频相关技术均有一定掌握和实践。同时，项目组成员均有一定开发经验，能够独立设计出完整的前端后端程序。

### 4.2.2劣势

作为初创项目组，我们的技术路线测试仅限于小范围封测，并没有进行公开测试，也没有在生产环境中进行测试。我们程序的稳定性有待检验

项目组成员作为本科生，知识储备较为有限，可能导致开发周期较长，程序也可能会出现不稳定的情况。

### 4.2.3机会

根据我们的市场调研，我们的调查对象有高达69.81%已经离不开社群分享类视频平台。而在这些人群中，又有仅仅8.33%的人没有遇到过视频配乐选取问题，有91.67%的人群在进行图片、视频配乐时遇到过各种各样的问题。我们发现为图片、视频进行人工智能配乐是一片尚未开发的蓝海。

### 4.2.4威胁

目前，各大科技公司都在加紧研发大模型解决一些实际问题。OpenAI、百度、谷歌、meta等公司由于掌握了较多算力，以及更多经验丰富的AI科学家。如果他们加入竞争，可能会开发出效果更好，运行更为稳定的产品，这会对我们产生较大影响。

## 4.3营销策略

目前，我们的产品主要面向短视频创作者和专业创作者，基于此，我们制定了以下的营销计划。

我们初步计划在短视频平台精准投放广告，向抖音等平台的视频创作者和小红书等平台的生活分享者定向推送我们的产品，从而对我们的产品进行营销。这种营销可以直达关键用户，使得我们的产品知名度得到提升。

在此之后，我们要尝试与部分视频网站的音乐区视频博主合作拍摄产品试用视频，以达到在专业制作人中的口碑推广以及树立在年轻人心中的形象。

## 4.4盈利模式

### 4.4.1贴片广告

我们初步计划在生成音乐的等待时间向用户推送视频广告以及贴片广告。用户观看广告后才能获得我们产品生成的配乐，进而通过广告盈利。

**4.4.2收费插件**

面向专业创作者，我们计划向他们提供定价合理的收费插件，在支付合理费用以后，创作者可以获取生成单个音乐的整个乐谱和MIDI文件，进而继续进行创作。同时，用户可以选择我们的会员制订阅服务，以更加优惠的价格享受我们的服务。

**4.4.3收费API以及技术支持**

在我们的技术经过多轮测试以及公众测试保证技术稳定性后，面向部分社交平台以及视频分享平台，我们可以与其达成技术合作，收取较为合理的接口使用费用，在收取费用之后与这些厂商达成战略合作关系，成为其技术提供商。

# 五、团队介绍

## 5.1导师介绍

**郭凯**：高级工程师，北京邮电大学物联网工程创新实践基地主任，研究方向为人工智能、物联网、电子商务、创新创业等。

**刘军：**博士，副教授，博士生导师，北京邮电大学数据科学中心主任，北京大数据协会常务理事。研究方向为人工智能及大数据。

**刘书昌：**毕业于中国科学院计算技术研究所，目前在北京邮电大学国际学院，主要从事数字音频信号处理、机器学习与人工智能、音频科技与数字媒体、创新创业教育等方面教学和科研工作。

## 5.2成员介绍

**陈雪松：**项目负责人

**徐天泽：**算法负责人

**李家俊：**算法工程师

**宁博睿：**商业负责人

**王乙竹：**前端工程师

**姚莘睿：**文职及美工

**刘恒余：**后端工程师

**王宇然：**后端工程师