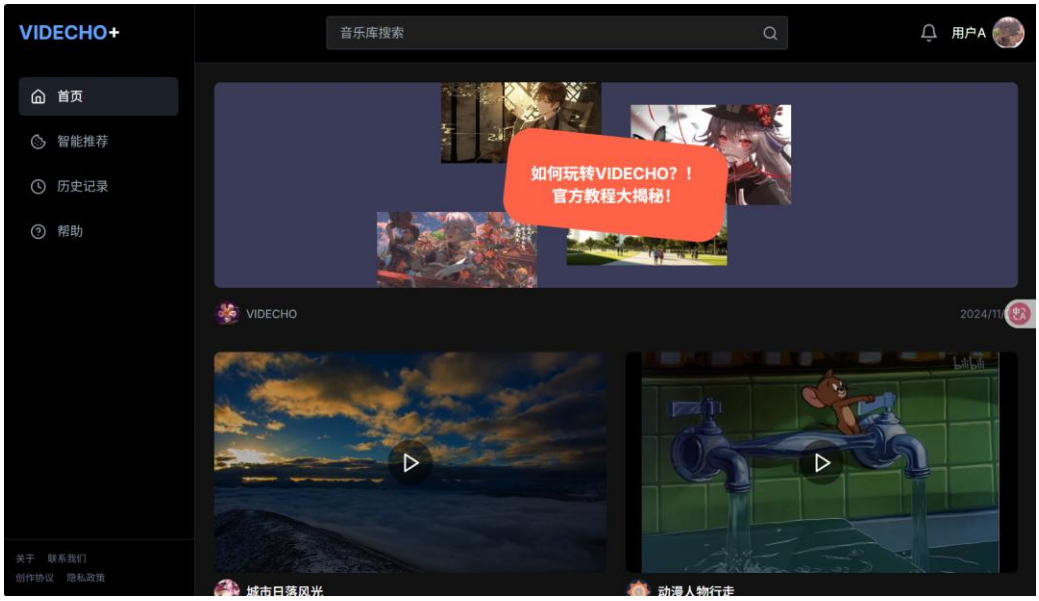
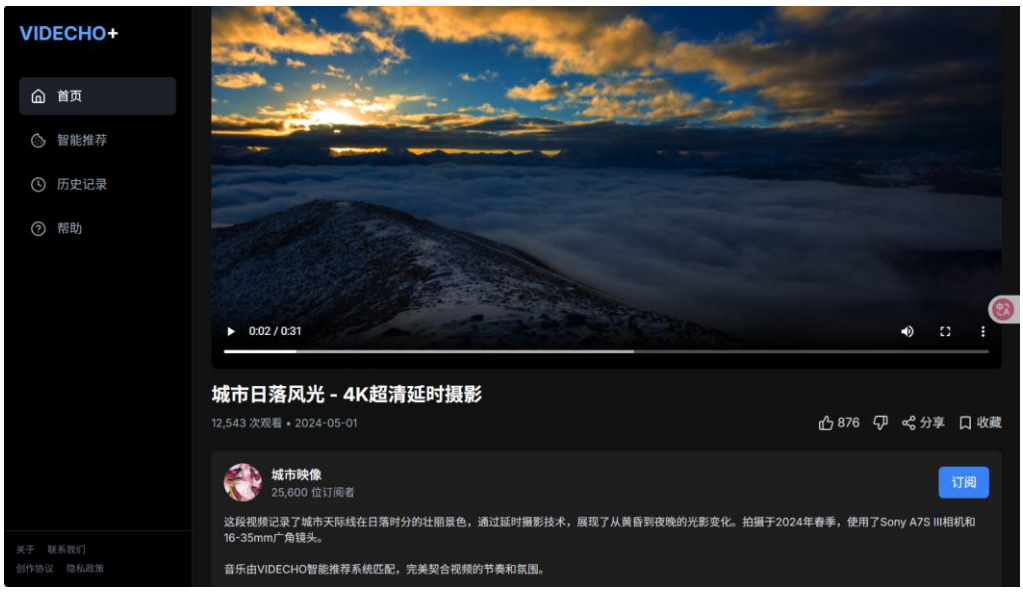
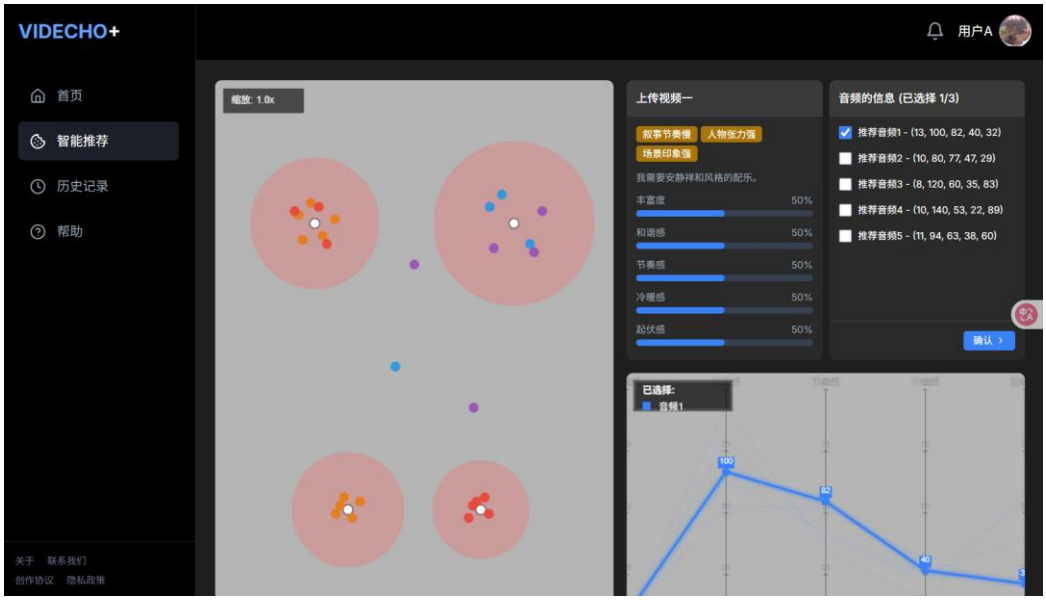
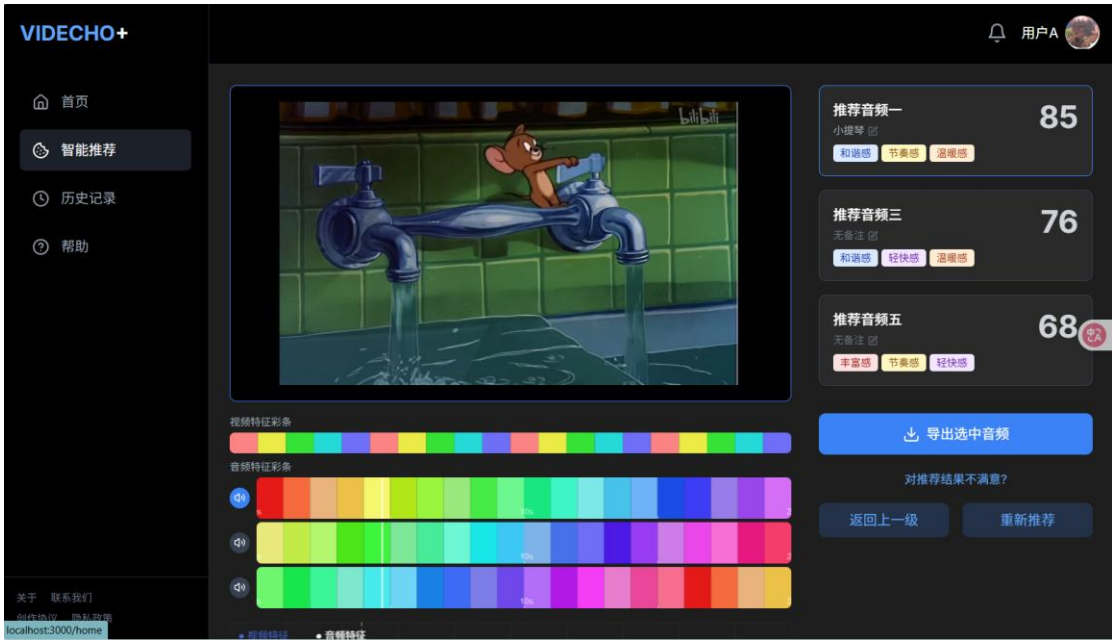
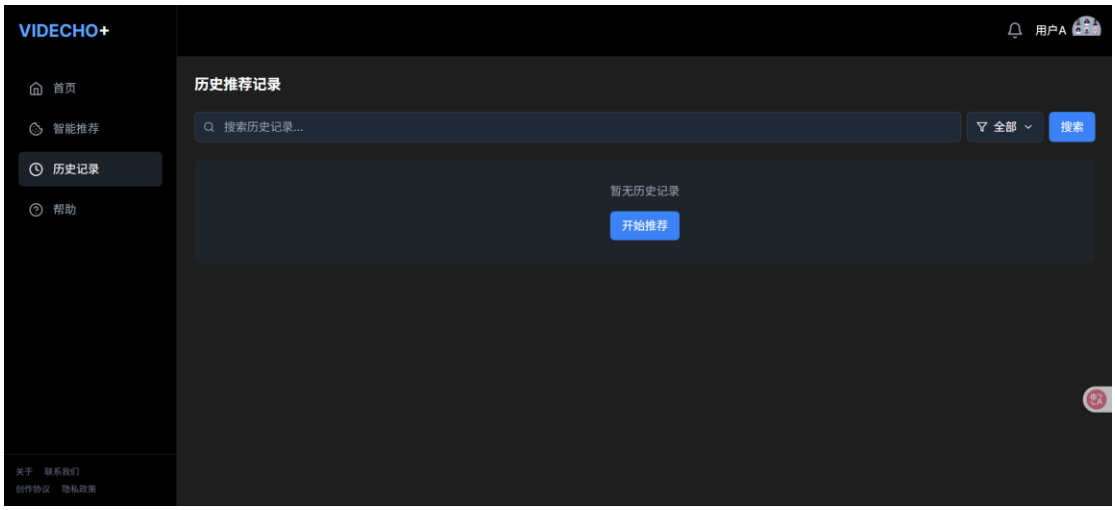
网页：https://kzmgzupjgllsaf6fge3a.lite.vusercontent.net/home

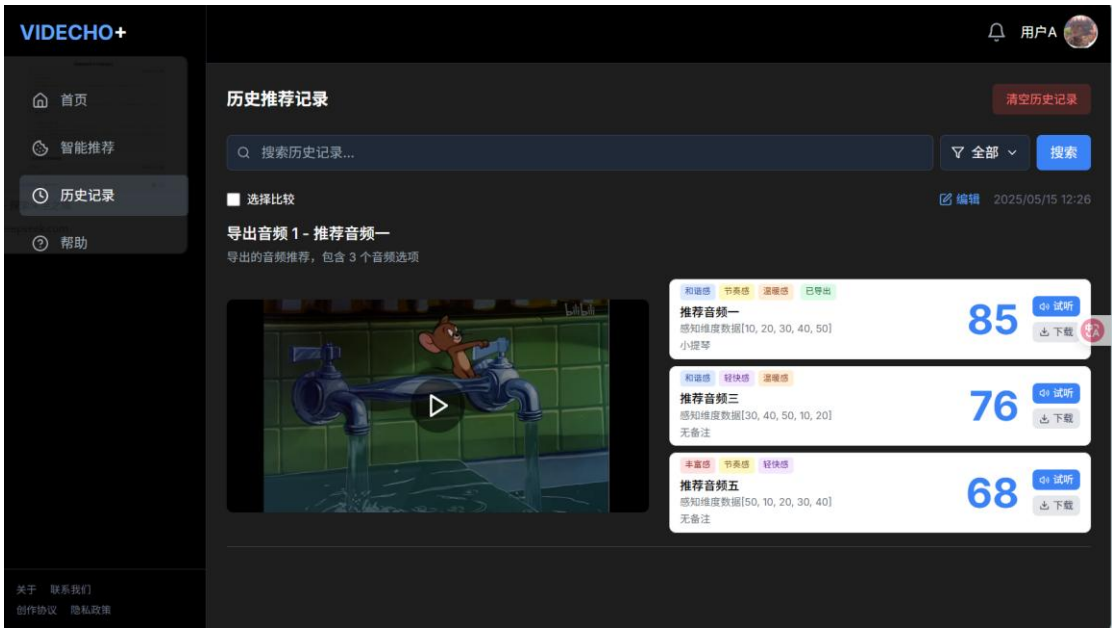





VIDECHO+
用户A
首页
智能推荐
历史记录
帮助
缩放: 1.0x
上传视频一
叙事节奏慢
人物张力强
场景印象强
我需要安静祥和风格的配乐。
丰富度 50%
和谐感 50%
节奏感 50%
冷暖感 50%
起伏感 50%
音频的信息 (已选择 1/3)
推荐音频1 - (13, 100, 82, 40, 32)
推荐音频2 - (10, 80, 77, 47, 29)
推荐音频3 - (8, 120, 60, 35, 83)
推荐音频4 - (10, 140, 53, 22, 89)
推荐音频5 - (11, 94, 63, 38, 60)
确认
已选择:
音频1
关于 联系我们
创作协议 隐私政策

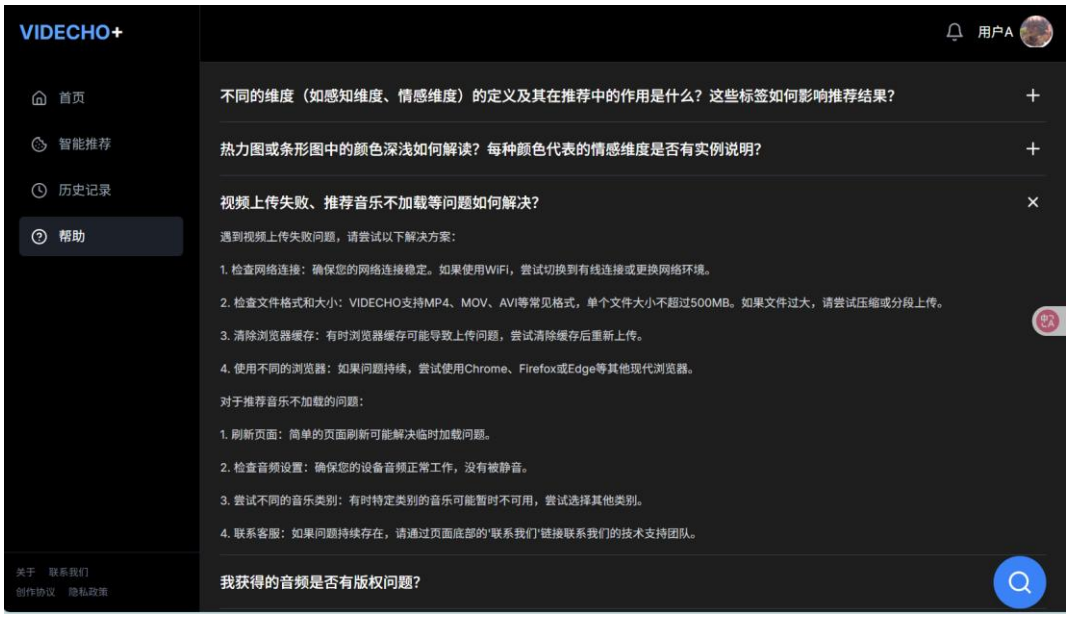

VIDECHO+
用户A
首页
智能推荐
历史记录
帮助
推荐音频一
小提琴
85
和谐感
节奏感
温暖感
推荐音频三
无备注
76
和谐感
轻快感
温暖感
推荐音频五
无备注
68
丰富感
节奏感
轻快感
视频特征彩条
音频特征彩条
导出选中音频
对推荐结果不满意?
返回上一级
重新推荐
关于 联系我们
localhost:3000/home


VIDECHO+
用户A
首页
智能推荐
历史记录
帮助
历史推荐记录
搜索历史记录...
全部
搜索
暂无历史记录
开始推荐
关于 联系我们
创作协议 隐私政策


VIDECHO+
用户A
首页
智能推荐
历史记录
帮助
历史推荐记录
清空历史记录
搜索历史记录...
全部
搜索
选择比较
编辑
2025/05/15 12:26
导出音频 1 - 推荐音频一
导出的音频推荐，包含 3 个音频选项
推荐音频一
感知维度数据[10, 20, 30, 40, 50]
小提琴
85
试听
下载
推荐音频三
感知维度数据[30, 40, 50, 10, 20]
无备注
76
试听
下载
推荐音频五
感知维度数据[50, 10, 20, 30, 40]
无备注
68
试听
下载
关于
联系我们
创作协议
隐私政策


VIDECHO+
用户A
首页
智能推荐
历史记录
帮助
历史推荐记录
清空历史记录
搜索历史记录...
全部
搜索
选择比较
编辑
2025/05/15 12:26
导出音频 1 - 推荐音频一
导出的音频推荐，包含 3 个音频选项
推荐音频一
感知维度数据[10, 20, 30, 40, 50]
小提琴
85
试听
下载
推荐音频三
感知维度数据[30, 40, 50, 10, 20]
无备注
76
试听
下载
推荐音频五
感知维度数据[50, 10, 20, 30, 40]
无备注
68
试听
下载
关于
联系我们
正在播放音频
ID: 5
0:07
-2:53
关闭
localhost:3000/home


VIDECHO+
用户A
首页
智能推荐
历史记录
帮助
不同的维度（如感知维度、情感维度）的定义及其在推荐中的作用是什么？这些标签如何影响推荐结果？
热力图或条形图中的颜色深浅如何解读？每种颜色代表的情感维度是否有实例说明？
视频上传失败、推荐音乐不加载等问题如何解决？
我获得的音频是否有版权问题？
关于
联系我们



模型介绍:

1. 视频特征提取模型

目的：

提取视频中与配乐相关的语义和感知特征，作为后续推荐和匹配模型的输入。

输入数据：

原始视频片段（目前支持 MP4）

提取特征维度：

目前视频特征包含如下 11 个维度:

人物张力、场景张力、叙事节奏、叙事内容、形象亲缘、动作能量、色彩轻重、色彩冷暖、动作幅度、画面唤醒度、画面效价

方法与工具：

视觉感知特征：通过预训练的 CNN 提取帧级动作与色彩特征。

剪辑节奏特征：通过帧间变化速率、运动矢量分析等方式获得动作节奏感。

张力与情绪特征：结合面部表情、镜头运动与场景检测，使用轻量 Transformer 进行时间建模。

输出格式：

每个视频最终编码为一个十一维的向量，作为推荐模型的输入。

（下图为特征预测值的 R 方）

	R2
人物张力	0.735714
场景张力	0.688761
叙事节奏	0.837422
叙事内容	0.829261
形象亲缘	0.776947
动作能量	0.751509
色彩轻重	0.467869
色彩冷暖	0.764292
动作幅度	0.410173
画面唤醒度	0.814311
画面效价	0.752924

2. 视频-音频推荐模型

目的：

根据视频特征向量，预测其匹配的音频特征向量，用于推荐最合适的配乐片段。

模型结构：

主干网络：多层全连接神经网络（MLP），将视频特征映射到音频空间。

多任务学习设计:

主任务：音频特征回归（9 维-8 维）

辅助任务：音频情绪感知回归（视频 VA-音频 VA）

损失函数：

回归损失：MSELoss，用于保持预测音频向量在向量空间中的位置与真实音频接近。

分类损失：CrossEntropyLoss，提升模型对音频情绪属性的理解。

总损失：主任务和辅助任务加权组合，权重通过超参数搜索确定最优并进行使用。

训练方法：

数据对构成：每个训练样本由一组（video_feat, audio_feat, audio_label）构成。

优化器：Adam / AdamW，带有学习率衰减机制。

超参数优化：采用 Optuna 框架对 hidden_dim、dropout、learning_rate、任务权重等进行调参，优化目标为 Recall@3。

批次训练：支持 GPU 多线程 DataLoader，mini-batch 大小一般为 32~64。

输出：

每个视频将被预测为一个 10 维的音频特征向量，用于匹配最接近的真实音频。

```
输入的视频特征：
人物张力: 60.7500
场景张力: 66.2000
叙事节奏: 54.0000
叙事内容: 56.8000
形象亲缘: -16.5000
动作能量: 66.8000
色彩轻重: 0.3800
色彩冷暖: 0.9700
动作幅度: 0.2900
画面唤醒度: 67.1400
画面效价: -29.2900

预测的音频特征：
Richness: -0.8372
Variation: 2.8622
Rhythm: -3.8342
Warmth: -5.1344
Harmony: -4.2335
brightness_score: 22.4444
Melody Trend Score: 50.9279
Rhythm speed score: 76.1797
音乐唤醒度: 56.2267
音乐效价: -20.4412
```

3. 音频匹配模型

目的：

将推荐出的音频特征向量与音频数据库中的所有音频进行相似度计算，返回 Top-K 最相似的配乐结果。

输入数据：

数据库：已提取的音频特征集 audio_features.csv，每个音频为一个 10 维向量

预测向量：由推荐模型输出的带权重的 10 维音频向量（用于调整权重匹配）

匹配算法：

支持**加权欧氏距离**和**余弦距离**两种相似度计算方式。用户可以切换匹配方式。

支持特征维度选择：如只使用唤醒度和效价（VA）进行匹配。

支持用户权重自定义：用户可对各维度设置重要性权重，影响最终匹配得分。

默认使用欧氏距离进行匹配。

匹配过程：

1. 载入数据库中所有音频特征向量
2. 对每个样本计算与预测向量的距离
3. 排序取前 Top-5 个最相似音频供用户筛选

输出：

返回最相似音频的名称、索引、匹配距离等结果，可用于配乐推荐界面展示。