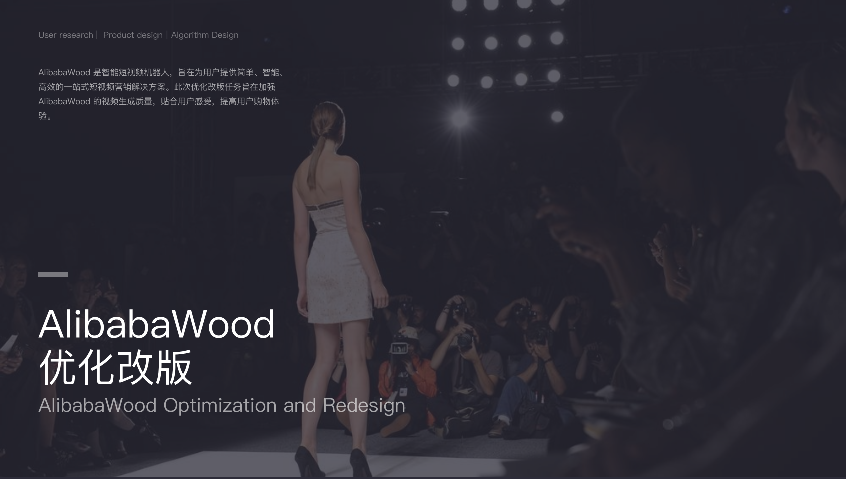


User research | Product design | Algorithm Design

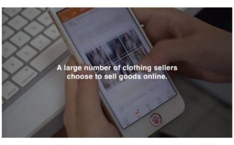
AlibabaWood 是智能短视频机器人，旨在为用户提供简单、智能、高效的一站式短视频营销解决方案。此次优化改版任务旨在加强 AlibabaWood 的视频生成质量，贴合用户感受，提高用户购物体验。

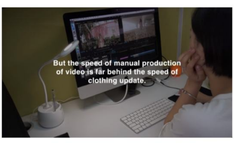
AlibabaWood
优化改版

AlibabaWood Optimization and Redesign

业务背景

AlibabaWood 是智能短视频机器人，旨在为用户提供简单、智能、高效的一站式短视频营销解决方案。它能自动获取并分析已有的淘宝、天猫等商品详情，组合生成商品短视频。


A large number of clothing sellers choose to sell goods online.


But the speed of manual production of video is far behind the speed of clothing update.

这对于需要大量**服装**新品上新的店家来说，可以有效地减少人员剪辑的劳动力，避免单一重复的疲劳工作。

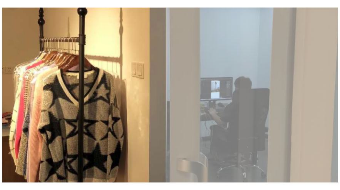
此次优化改版任务，旨在**加强 AlibabaWood 的视频生成质量，贴合用户感受，提高用户购物体验。**

前期调研

用户需求挖掘

User Demand Generation

- 对未使用者进行调研



地 / 四季青线服装卖场　图 / 黄小倚

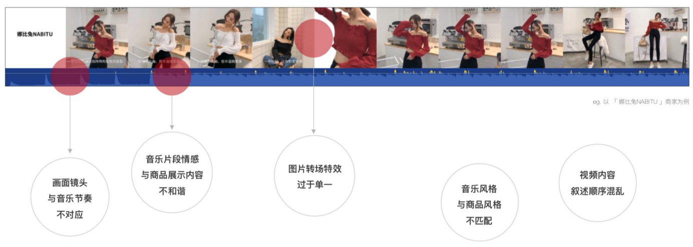
采访对象：**经营线上淘宝店且未使用过AlibabaWood的卖家**
店家数量：40+
基本特征：**周上新量大(>100)**
① 有空的时候会自己拍摄视频(不剪辑)
② 忙碌的时候只拍照片不拍视频(无视频)
③ 需要聘请视频剪辑员进行素材整合
④ 视频制作人力、时间消耗量大，很难审核质量

使用意愿：**愿意使用自动生成工具提高制作效率**

线下调研中，我们看到剪辑师正在商店里屋进行视频剪辑工作，也对其进行了采访。**剪辑师表示，会考虑使用自动生成工具来辅助视频制作，以提高效率。**

- 对已使用者进行调研

针对存在问题进行了相互验证与整合，得出如下结论：


娜比兔NABITU
画面镜头
与音乐节奏
不对应
音乐片段情感
与商品展示内容
不和谐
图片转场特效
过于单一
音乐风格
与商品风格
不匹配
视频内容
叙述顺序混乱

eg. 以 「娜比兔NABITU」商家为例

- 寻找学术支持

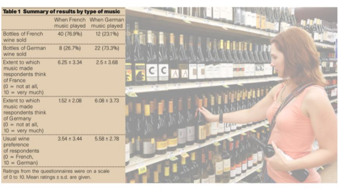


Table 1 Summary of results by type of music

	When French music played	When German music played
Bottles of French wine sold	40 (76.9%)	12 (23.1%)
Bottles of German wine sold	8 (26.7%)	22 (73.3%)
Extent to which music made respondents think of France (0 = not at all, 10 = very much)	6.25 ± 3.34	2.5 ± 3.68
Extent to which music made respondents think of Germany (0 = not at all, 10 = very much)	1.52 ± 2.08	6.08 ± 3.73
Usual wine preference of respondents (0 = French, 10 = German)	3.54 ± 3.44	5.58 ± 2.78

Ratings from the questionnaires were on a scale of 0 to 10. Mean ratings ± s.d. are given.

North, Adrian C., David J. Hargreaves, and Jennifer McKendrick. "In-store music affects product choice." Nature (1997): 132.

正确关联的音乐 对比 错误关联的音乐
可以使红酒销量销量提升 2.75–5 倍

因此，**在线上购物环境中，不仅是商品展示的视觉效果，背景音乐的听觉感受也能给人带来较大的情绪影响，从而影响购物决策。**

需求剖析

Demand analysis

基于以上旧版本存在痛点及用户需求，总结出以下四个改进方向。
但考虑到方案实现难度以及业务持续时间限制，最终确定：**新版本先完成「音乐推荐」和「音乐/视频转场」两个方向的改进。**

音乐推荐
- **从有版权音乐库中推荐与商品风格相符的音乐**

音乐/视频转场
- **找出音乐中的转场点**
- **增加视频转场特效**

音乐结构
- 不同音乐片段存在不同的情感表现，找出更适合与商品属性适配的音乐片段

视频内容
- 商品介绍内容需要存在故事性，顺序需要调整
- eg. 商品整体展示–局部展示–细节展示–商品说明–商家LOGO

视听融合可以实现短视频的音画联觉，呈现统一的视听感受。

视觉效果
- 画面流畅
- 图片/镜头切换效果和谐

→

视听融合
- 商品和音乐搭配和谐
- 图片/镜头切换符合音乐节奏转折

←

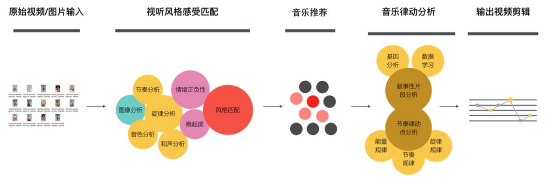
听觉感受
- 符合用户喜好
- 歌曲促进购买情绪

设计

设计框架

Design Framework

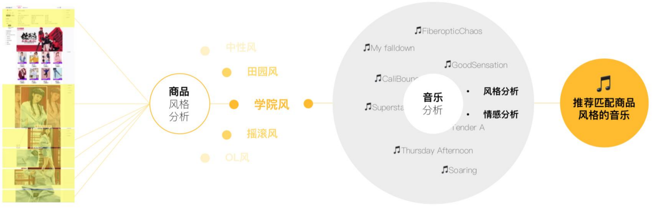
建立服装风格与音乐属性(包括情感和风格)的对应关系 ， 根据热度信息推荐匹配商品风格的背景音乐；并通过对背景音乐的特征分析和理解，进行基于音乐节奏的转场点识别和推荐，在转场点之间插入视频图片；实现短视频的音画联觉，呈现统一的视听感受。


原始视频/图片输入
视听风格感受匹配
音乐推荐
音乐律动分析
输出视频剪辑
节奏分析
情绪正负性
图像分析
旋律分析
风格匹配
唤起度
音色分析
和声分析
基因分析
数据学习
叙事性片段分析
节奏律动点分析
能量规律
旋律规律
节奏规律

设计方法 Design Approach

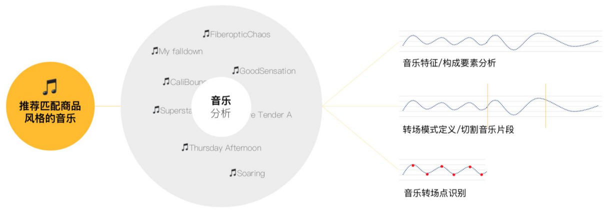
商品音乐推荐

基于服装风格的音乐推荐方案的基本思路是用音乐属性信息作为桥梁，将不同的服装风格标签与不同的音乐对应起来，为服装卖家推荐符合其风格的背景音乐。

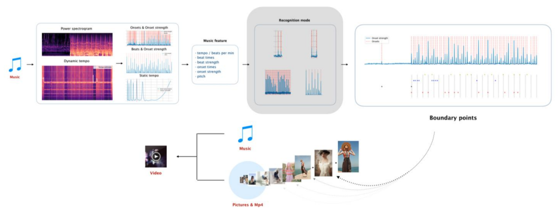

商品
风格
分析
中性风
田园风
学院风
摇滚风
OL风
FiberopticChaos
My falldown
GoodSensation
CaliBounce
Supersta
Tender A
Thursday Afternoon
Soaring
音乐
分析
风格分析
情感分析
推荐匹配商品
风格的音乐

音乐转场点识别 ★

基于音乐特征的转场点识别与推荐方案的基本思路是对推荐的音乐进行特征分析，根据特征模式识别，确定音乐转折点，从而推进适合视频转场的时间点。这样的自动化工具可以极大地提高音乐和视觉元素的对齐效果，达到视听融合的自动化剪辑目的。


推荐匹配商品
风格的音乐
FiberopticChaos
My falldown
GoodSensation
CaliBounce
Supersta
Tender A
Thursday Afternoon
Soaring
音乐
分析
音乐特征/构成要素分析
转场模式定义/切割音乐片段
音乐转场点识别

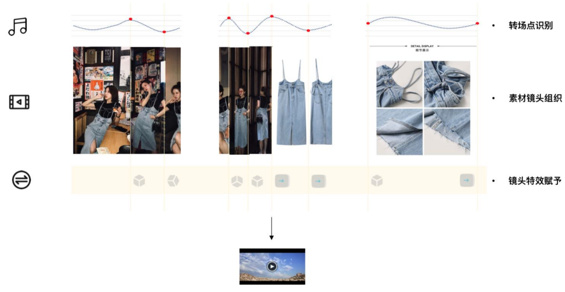
基于音乐特征的转场点识别与推荐方法总共分为3大模块——<音乐特征提取>，<转场模式定义>，<转场点识别与优化排序>。


Music
Power spectrogram
Onsets & Onset strength
Dynamic tempo
Beats & Onset strength
Static tempo
Music feature
tempo / beats per min
beat times
beat strength
onset times
onset strength
pitch
Recognition mode
Boundary points
Music
Video
Pictures & Mp4

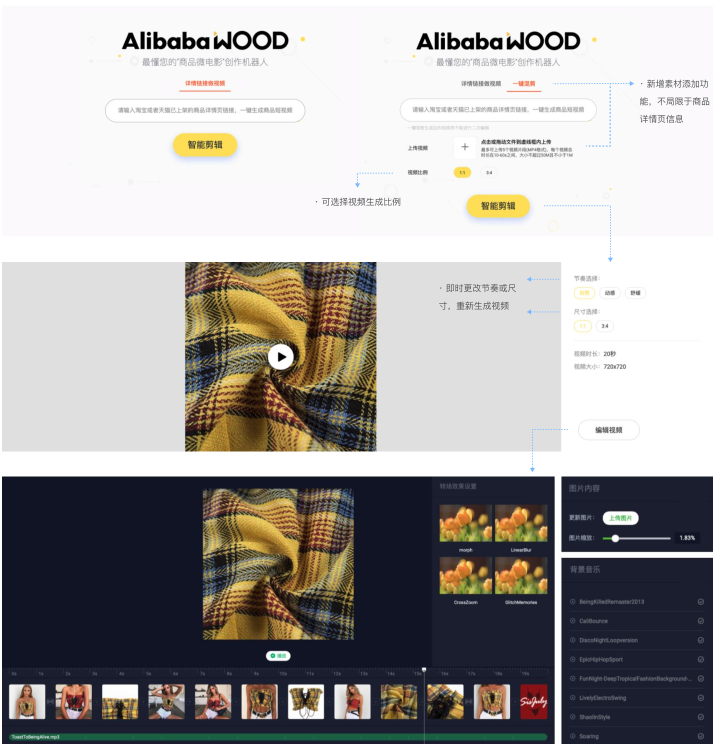
- **转场模式定义**

① 邀请随机大众(音乐专业与非专业人员)，试听音乐。
② 记录被试听歌时的表情、语言、行为变化，并记录出现表现时的音乐时间点。
③ 通过用户听歌后的主观感受访谈与数据分析(音乐点附近特征分析)，得出常见转场模式。

模块整合 ★


DETAIL DISPLAY
细节展示
转场点识别
素材镜头组织
镜头特效赋予

设计效果 Design Effect


AlibabaWOOD
最懂您的"商品微电影"创作机器人
详情链接做视频
一键混剪
智能剪辑
上传视频
视频比例
1:1
3:4
节奏选择：
自然
动感
舒缓
尺寸选择：
编辑视频
转场效果设置
morph
LinearBlur
CrossZoom
GlitchMemories
播放
图片内容
更新图片：
上传图片
图片缩放：
1.83%
背景音乐
BeingKilledRemaster2013
CaliBounce
DiscoNightLoopversion
EpicHipHopSport
FunNight-DeepTropicalFashionBackground-...
LivelyElectroSwing
ShaolinStyle
Soaring
ToastToBeingAlive.mp3

请输入淘宝或者天猫已上架的商品详情页链接，一键生成商品短视频

一键混剪生成后的视频将不能进行二次编辑

点击或拖动文件到虚线框内上传
最多可上传5个视频片段(MP4格式)，每个视频总时长在10-60s之间，大小不超过50M且不小于1M

视频时长：20秒
视频大小：720x720

· 新增素材添加功能，不局限于商品详情页信息

· 可选择视频生成比例

· 即时更改节奏或尺寸，重新生成视频

用户体验

操作实验 Operation Experiment

此次用户体验实验采用的是AlibabaWood上线前的demo版本 —— ApVideor工具。(ApVideor : A Video Automaker for Apparels)

实验方法：线下视频制作 (手动剪辑 + 使用系统智能生成视频)

被试人数：17位年龄在22～24岁视频编辑员 (9位女性和8位男性，视频编辑员定义为具有视频编辑工作经验的人)

邀约形式：通过社交平台、校内论坛传播信息

实验大纲：A. 第一步，每个参与者分别为两件商品手动创建视频(每个视频的长度控制不到30秒)，允许使用他们喜欢的任何视频编辑软件。在创建之后，参与者需要进行3项主观评估：1.视频编辑的总难度级别，范围为0到5(0表示非常容易，5表示非常困难)；2.寻找背景音乐；3.合成音乐和视觉效果。
B. 第二步，参与者需要使用 ApVideor 工具自动创建视频并评估难度级别。

我们将记录报告的评级和被试完成每项任务所花费的时间。

观看实验 Watching Test

我们将对上一步用户体验实验中通过两种方式产出的视频进行用户观看测试，以验证视频质量。

实验方法：线上视频观看 + 打分

被试人数：70名 (41名女性和29名男性，参与用户研究的受试者被排除在外)

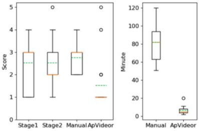
邀约形式：通过社交平台、校内论坛传播信息

实验大纲：A. 测试者需要评估音乐与服装风格的匹配程度以及视觉过渡与音乐的匹配程度，范围为1到5（5表示完全匹配，4表示适度匹配，3表示可接受，2表示不那么令人愉快，1表示完全不合适）。

B. 测试者对手动创建的视频与ApVideor创建的视频进行总体效果感受比较。

数据分析 Data Analysis

- **使用ApVideor的创建难度水平显着降低**（t = 3.8996，p = 0.0003626）
- 同时，**手动剪辑所花费的时间(2个视频)也比使用ApVideor自动生成视频花费的时间长得多**（t = 13.089，p = 1.271×10−10）

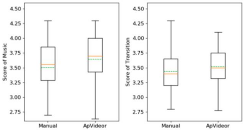


橙色实线表示中值，绿色点线表示平均值

左图：视频创建难度(实验中的4项评估)

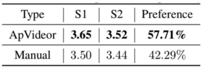
右图：两个任务所需时间

- 经T检验验证，**ApVideor可以找到与服装风格匹配的音乐并产生视觉切换效果，视频质量不差于有视频编辑工作经验的人制作的视频**



上图显示了两个任务产出的视频分别由测试者评分的两个分数分布。由于这两个分布之间没有显着差异(左图)，我们使用two-one-sided t-tests (TOST)来执行等效性检验(右图)。

- **ApVideor 创建的视频的评级高于手动创建的视频，无论是背景音乐还是视觉过渡**
- 同时，我们可以从同一张表中看到，**测试者也更偏爱ApVideor 创建的视频**



Type	S1	S2	Preference
ApVideor	**3.65**	**3.52**	**57.71%**
Manual	3.50	3.44	42.29%

表格显示背景音乐(S1)和视觉过渡(S2)的平均分数，以及受试者报告的偏好。

优化结果 Optimized Result

购物转化率

↑ 10–30%