

bilibili

前端在数字人创作工具中的应用实践

邓斌 Bean



必剪 Studio



必剪 Studio

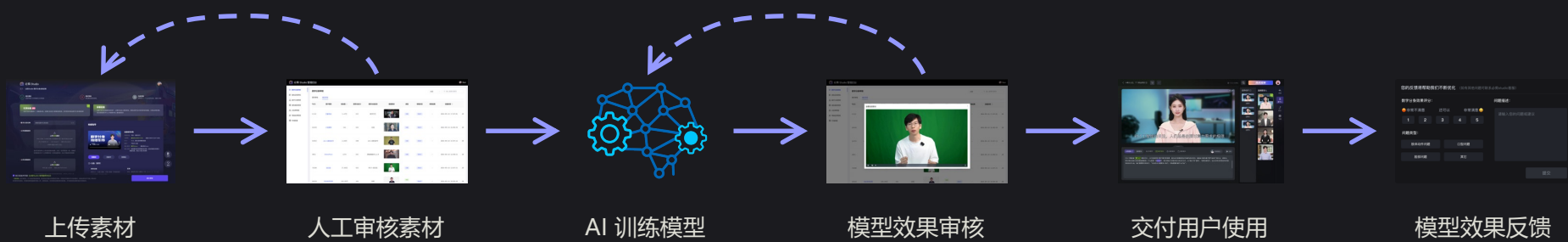
必剪 Studio 简介

- 基于 **音色克隆** 与 **数字分身** 技术的智能创作平台
- **音色克隆**: 复刻个人音色, 通过文本生成音频
- **数字分身**: 通过音频驱动嘴型生成口播视频
- 目标: 降低视频创作门槛, 提升视频创作效率



必剪 Studio 简介

- AI 模型线上训练工业化流程



绿幕抠像

- 什么是绿幕抠像 (Chroma key)
 - 去除绿幕背景替换为其他背景图/视频
 - 通常使用绿幕或蓝幕
 - 广泛应用于电影、电视、视频制作等领域
- 为什么要支持绿幕抠像
 - 训练绿幕的数字分身，可实现一次训练，在合成时任意换背景



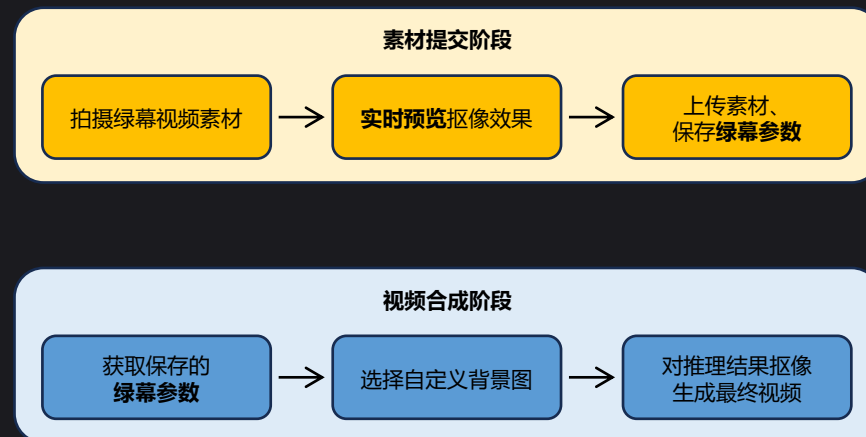
绿幕抠像

● 绿幕抠像的挑战

- 纯色绿幕素材：剪辑要求高，很多素材不达标
- 实景绿幕素材：拍摄光照、绿幕颜色等不统一
- 审核人力有限，如何高效设置绿幕抠像的参数

● 解决方案：

- 将抠像参数设置从素材审核阶段前置到素材提交阶段
- 素材提交时让用户实时预览抠像效果并保存参数



绿幕抠像

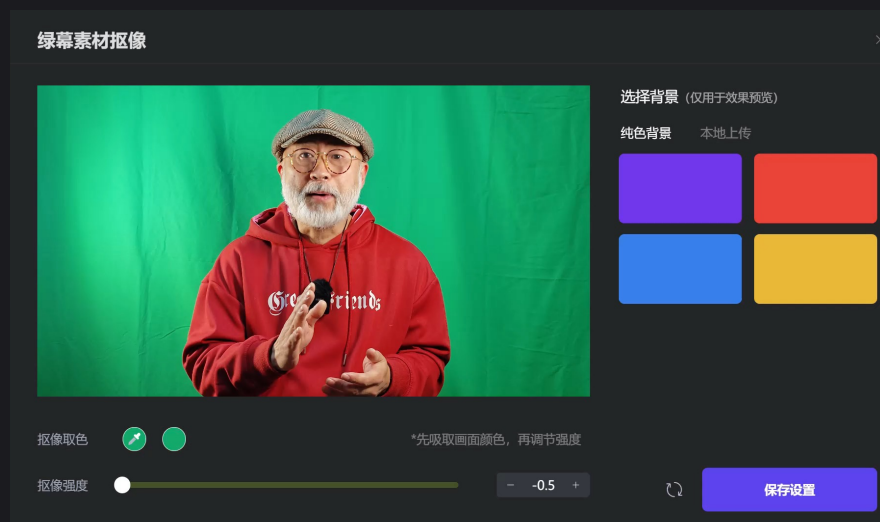
- 实时抠像预览实现原理

- 逐帧获取图像
- 利用 WebGLShader 计算每个像素点的透明度
- 处理后的图片逐帧绘制到 Canvas 上

- 用户参数

- 绿幕色值 (自动获取 + 手动吸色)
- 相似度 (手动拖拽)

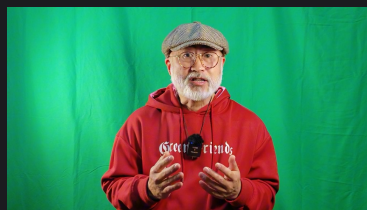
- 处理速度: 2ms/帧



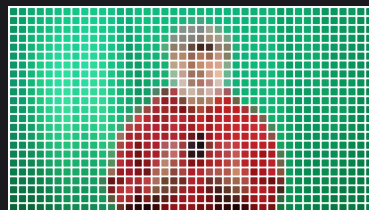
前端实时抠像效果演示

绿幕抠像

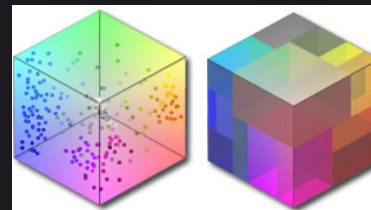
- 如何自动获取绿幕颜色



截取视频首帧图像



获得每个像素点的色值



色彩量化 (中值切割)



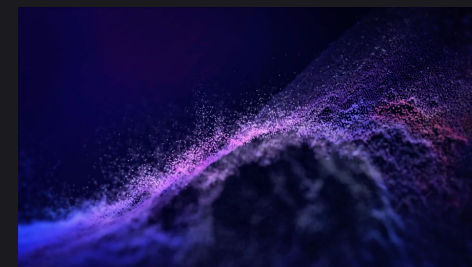
Key Color

音视频合成

- 基于 WebCodecs + FFmpeg.wasm 双引擎封装
 - <https://github.com/hughfenghen/WebAV> @风痕
 - <https://github.com/ffmpegwasm/ffmpeg.wasm>
- 核心能力
 - 音视频拆分、合成、绿幕抠像、画中画、水印、字幕、转场特效、元数据编辑等



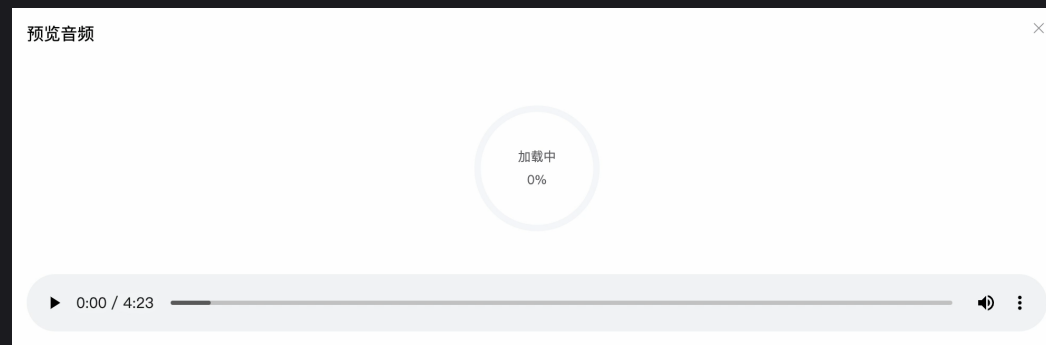
画中画、字幕、水印效果



转场效果 (FFmpeg xfade)

音频波形可视化

- 后台审核音频可视化波形图
 - 方便定位特定位置
 - 直观看音量过高或过低
 - 快速分辨重复拼接的音频
- 借助 `wavesurfer.js` 库实现
 - <https://github.com/katspaugh/wavesurfer.js>



音频波形可视化效果演示

音频在线转码

- 解决 ALAC 编码音频文件无法直接在 Chrome 浏览器播放的问题
- iPhone 录音机导出音频默认为 ALAC 编码的 .m4a 文件
- 在 onerror 事件中加载 FFmpeg.wasm 在线转码后替换原地址



Apple Lossless Audio Codec
苹果无损音频压缩编码



SSML 可视化编辑器

- SSML (语音合成标记语言) 是 [W3C](#) 语音接口框架的一部分
- 基于 XML, 支持 break (停顿)、phoneme (发音)、say-as (读法)、sub (替换)、prosody (韵律) 等标签
- 主要解决 TTS 中多音字、停顿、发音、读法等定制化需求
- 将 SSML 的编辑可视化, 让普通用户拥有调整能力

```
<small>xml</small>
<code>< speak >
  今天天气< break time="2.5s" ></ break >很好。
  我< phoneme alphabet="py" ph="bei4" >背</ phoneme >着爸爸去踢球了。
  < say-as interpret-as="number" >1234567</ say-as >
  < say-as interpret-as="digits" >7654321</ say-as >
  < sub alias="Speech Synthesis Markup Language" >SSML</ sub >
</ speak ></code>
```

SSML (Speech Synthesis Markup Language) 示例

SSML 可视化编辑器

The screenshot displays the SSML visualization editor interface. At the top, there is a navigation bar with several tabs: "文字驱动" (Text-driven), "录音驱动" (Recording-driven), "多音字" (Polyphonic characters), "数字读法" (Number reading), "发音替换" (Pronunciation replacement), and "插入停顿" (Insert pause). The "文字驱动" tab is currently selected. To the right of the navigation bar, there is a user profile icon labeled "Bean" and a "试听" (Preview) button. The main area of the editor contains a text input field with the following text: "bilibili 创立于 2009 年，是国内知名的视频弹幕网站，也被称为“B 站”。2023 年全年，B 站日活用户平均超过 9800 万。". At the bottom right of the editor, there are two buttons: "全屏编辑" (Full screen edit) and "72 / 5000" (likely indicating character count or position).

SSML 可视化编辑器效果演示

Thanks!



哔哩哔哩技术

微信扫描二维码，关注我的公众号