大家好，我是**北小菜**，今天介绍视频行为分析系统v3.46版本

**\* 没有看过此前任何相关视频的用户，也可以从本视频直接开始**

**\* v3.46最大变化：引入deepseek级多模态视觉大模型MiniCPM**

**v3安装包下载地址：**

<https://gitee.com/Vanishi/BXC_VideoAnalyzer_v3>

1. **首先介绍v3.46主要变化：**
   * 1. v3.46之前已经发布了3.0，3.1，3.2，3.3，3.40，3.41，3.42，3.43，3.44，3.45 等10个版本（了解不同的v3版本，可以参考README）
     2. v3.46 是在v3.45的基础上，删除了TensorRT推理引擎的功能实现，之所以删除TensorRT，是因为很多用户的电脑没有配置满足TensorRT的N卡，无法正常使用v3.45。

除此之外，v3.46最大的特点是引入了多模态视觉大模型，具体是基于llama.cpp推理引擎推理MiniCPM大模型，该组合最大的优势在于支持绝大部分普通设备， 普通用户基于v3.46的Windows软件安装包，可以实现一键启动，真正实现0门槛体验视觉多模态大模型。v3.46有关推理MiniCPM大模型的代码也是比较有参考学习价值！



1. **v3.46支持哪些硬件：**
   * 1. v3.46支持两种小模型推理引擎，分别是：OpenVINO，OnnxRuntime
     2. v3.46支持一种大模型推理引擎，是llama.cpp

OpenVINO可以支持x86/arm架构，可以支持Windows/Linux等系统。 但是仅可以在包含核显的6代以上英特尔处理器上加速，非英特尔核显处理器虽然可以运行，但是无法加速。

OnnxRuntime可以支持x86/arm架构，可以支持Windows/Linux等系统。但是v3整个系列内置的OnnxRuntime均未设置加速后端，如需加速，需要用户自行二次开发。

llama.cpp可以支持x86/arm架构，可以支持Windows/Linux等系统。对硬件的要求极低，例如运行v3.46内置的MiniCPM视觉大模型，只需要满足内存大于等于8G即可。

1. **v3.46如何编译：**

接下来会演示如何在x86/Windows系统上编译：编译后台管理Admin，编译分析器Analyzer，编译流媒体服务器MediaServer

**特别注意：如果需要在非x86/Windows系统编译v3.46，要特别注意分析器模块，Analyzer/CMakeLists依赖的第三方库需要自行准备**