



北京异构智能科技有限公司

# 视频分析平台

产品说明书

# 目录

产品描述.....	3
目标用户.....	3
解决方案.....	3
产品特点.....	3
视频结构化流程.....	3
产品说明.....	4
视频捕获.....	4
视频处理.....	4
NovuTensor .....	5
视频分析 .....	5
模型加载器 .....	6
任务管理系统.....	6

## 产品描述

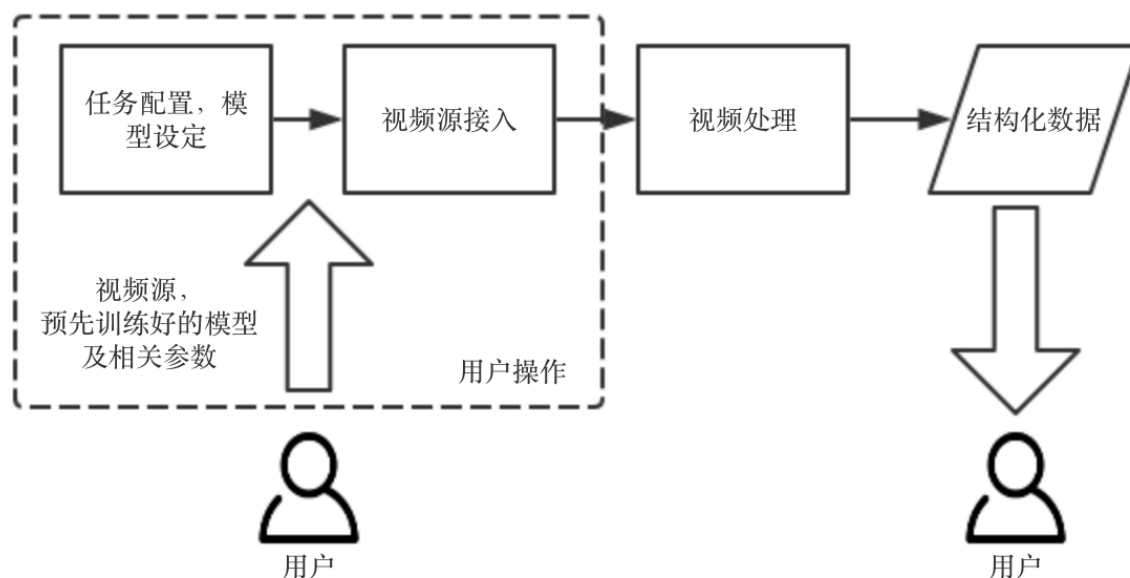
本产品为基于深度学习的视频分析平台，可根据不同分析需求动态加载深度神经网络模型，灵活配置，实现了简单快速的多视频分析处理。

## 目标用户

需要简单快速实现视频分析的企业级用户

## 解决方案

提供从视频源接入，到结构化数据输出的一整套视频分析解决方案。

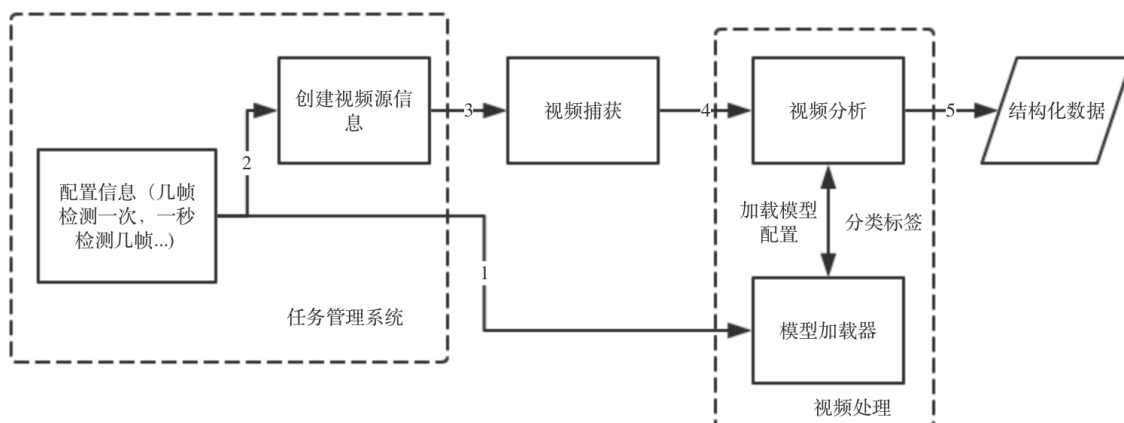


## 产品特点

1. 多路视频源并行处理
2. 神经网络模型“即插即用”，复数加载
3. 视频分析任务配置化
4. 多路推算

## 视频结构化流程

1. 加载配置信息与模型
2. 创建视频源信息
3. 视频捕获
4. 视频处理
5. 生成结构化数据

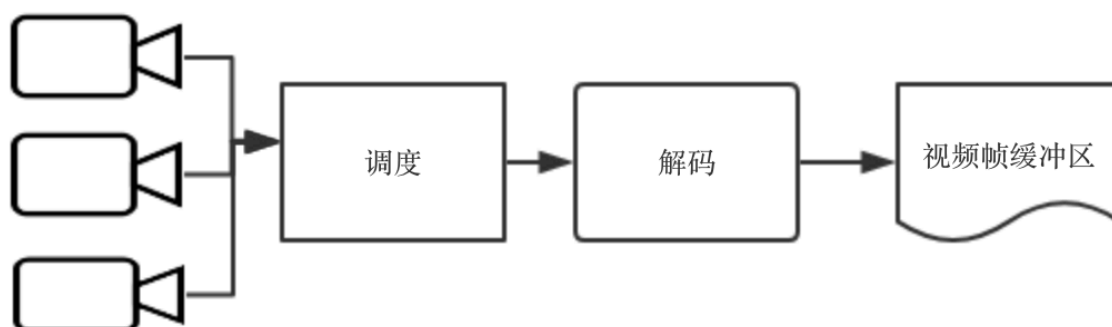


## 产品说明

本产品由三大模块组成：视频捕获，视频处理，以及任务管理系统。

### 视频捕获

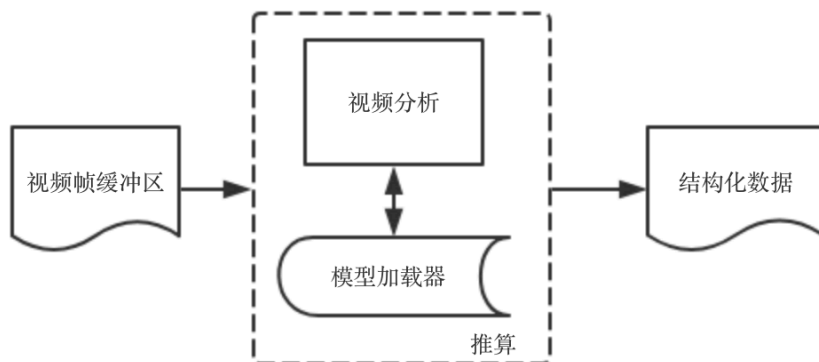
视频捕获主要实现了并行视频流的调度处理以及视频解码。工作流如下图所示：



- 可支持输入视频源有 USB 摄像头，CSI 摄像头，RTSP 视频流，视频文件
- 视频调度的作用是对实时新增或者删除的视频源进行处理
- 视频解码过程中会将视频源的信息附加到每帧中
- 最终将视频源逐帧分配到视频处理层

### 视频处理

视频处理为本产品的核心模块，基于 Novumind 的深度学习推演加速芯片 NovuTensor，结合深度神经网络模型，对视频帧进行检测，分类，特征提取，位置追踪，最终生成结构化数据。工作流如下图所示：



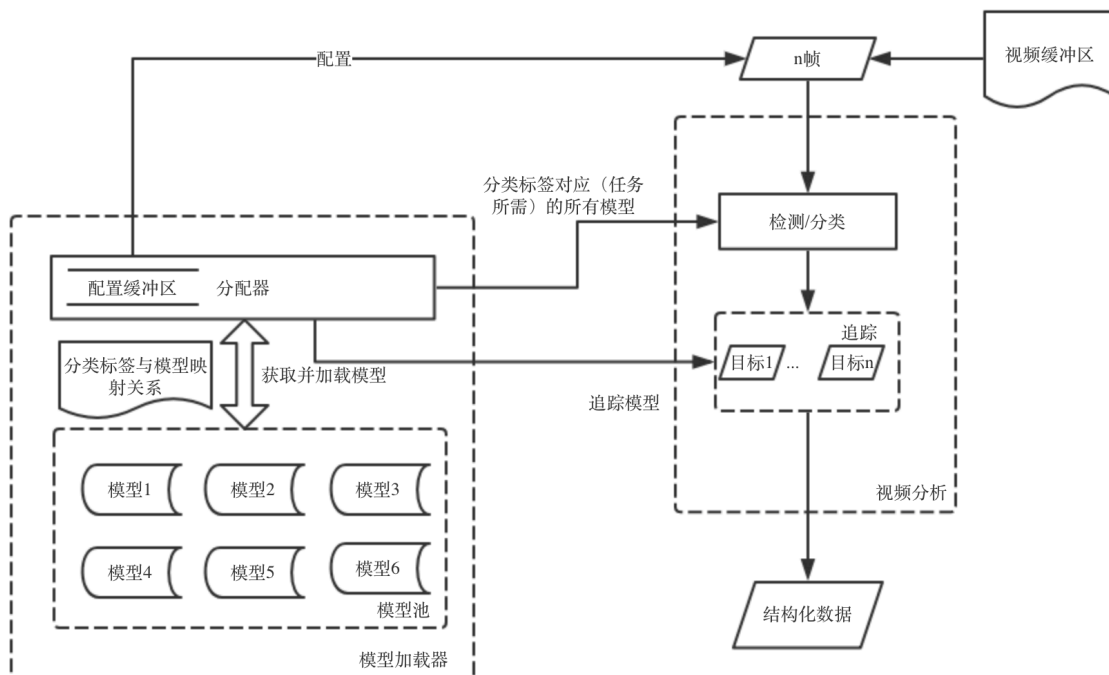
## NovuTensor

Novumind 公司设计研发的深度学习推理加速芯片，可提供高效地卷积神经网络推理，最快可以处理每秒 15 兆的运算。在 NovuTensor 强大的算力支持下，能够并行实时处理多路视频，而且能够同时加载多个模型，处理不同的分析任务。根据处理的模型大小以及分辨率的不同，同时加载的模型数量以及视频路数会有不同。例如一张 NovuTensor 芯片可以以约 240fps 运行 yolov2 (416\*416)，或以 30fps 8 路并行运行。一张 PCI-e 卡可同时运行 4 张 NovuTensor。

## 视频分析

工作流程（见下图）：

1. 模型加载器根据任务获取所需模型
2. 根据配置从视频缓冲区选取 n 帧视频
3. 对 n 帧视频进行检测分类处理，并追踪目标，生成结构化数据



## 模型加载器

模型加载器由分配器，模型池组成。

分配器根据不同的分类标签以及任务，从模型池中获取所需模型（多个独立模型/一个复合模型），并创建模型资源流，每个资源流加载模型服务于不同的分析任务。分配器负责管理模型的生命周期，包括模型的获取，加载，以及卸载。

模型池作为一个磁盘空间，用于放置训练好的各种模型。配置信息、模型以及分类标签与模型的映射关系可通过任务管理系统进行管理。

## 任务管理系统

负责视频源管理，模型管理，以及配置管理。

主要功能有：

- 视频/视频源管理
  - 添加视频
    - 上传多个视频文件
    - 批量添加视频流地址
  - 删除视频
- 模型管理
  - 分类标签与模型映射关系管理
  - 模型的添加/删除
- 配置管理（可配置 **fps**，几路视频并行处理等）
  - 添加，删除配置项