



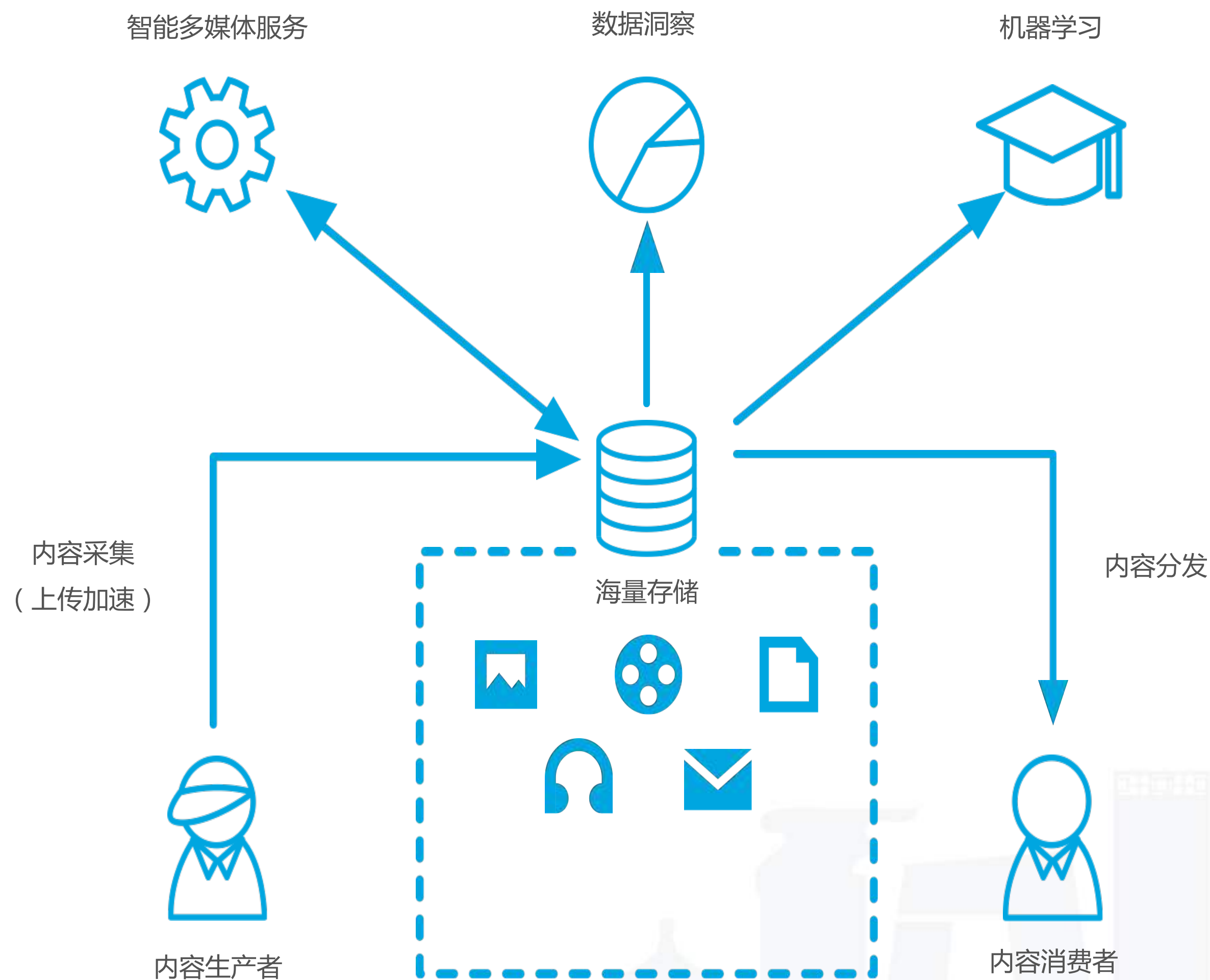
QCon 全球软件开发大会
INTERNATIONAL SOFTWARE
DEVELOPMENT CONFERENCE

BEIJING 2018

《基于深度学习的视频结构化实践》

七牛云 AI实验室首席架构师 / 姚唐仁

关于七牛云



- 国内领先的云计算厂商

- 围绕海量数据提供创新的云服务，帮助客户缩短想法到产品的距离
- 创立6年，每年超过300%的业绩增长
- 已完成5轮融资，累计超过20亿

- 长期服务70多万企业用户和开发者

- 文件数超过2000亿，每日新增文件20亿
- 覆盖全球300个节点
- 覆盖金融、公安、广电媒体、互联网等行业

海量富媒体数据的业务布局

视频+智能

视觉智能 Vision Intelligence



内容审核



人脸识别



视频分析



深度学习



多媒体API

数据智能 Data Intelligence



数据分析决策



理解机器语言和情绪



洞察未知数据



直播云



点播云



实时音视频互动



播放器

弹性计算



容器云



云主机

海量存储



智能网络

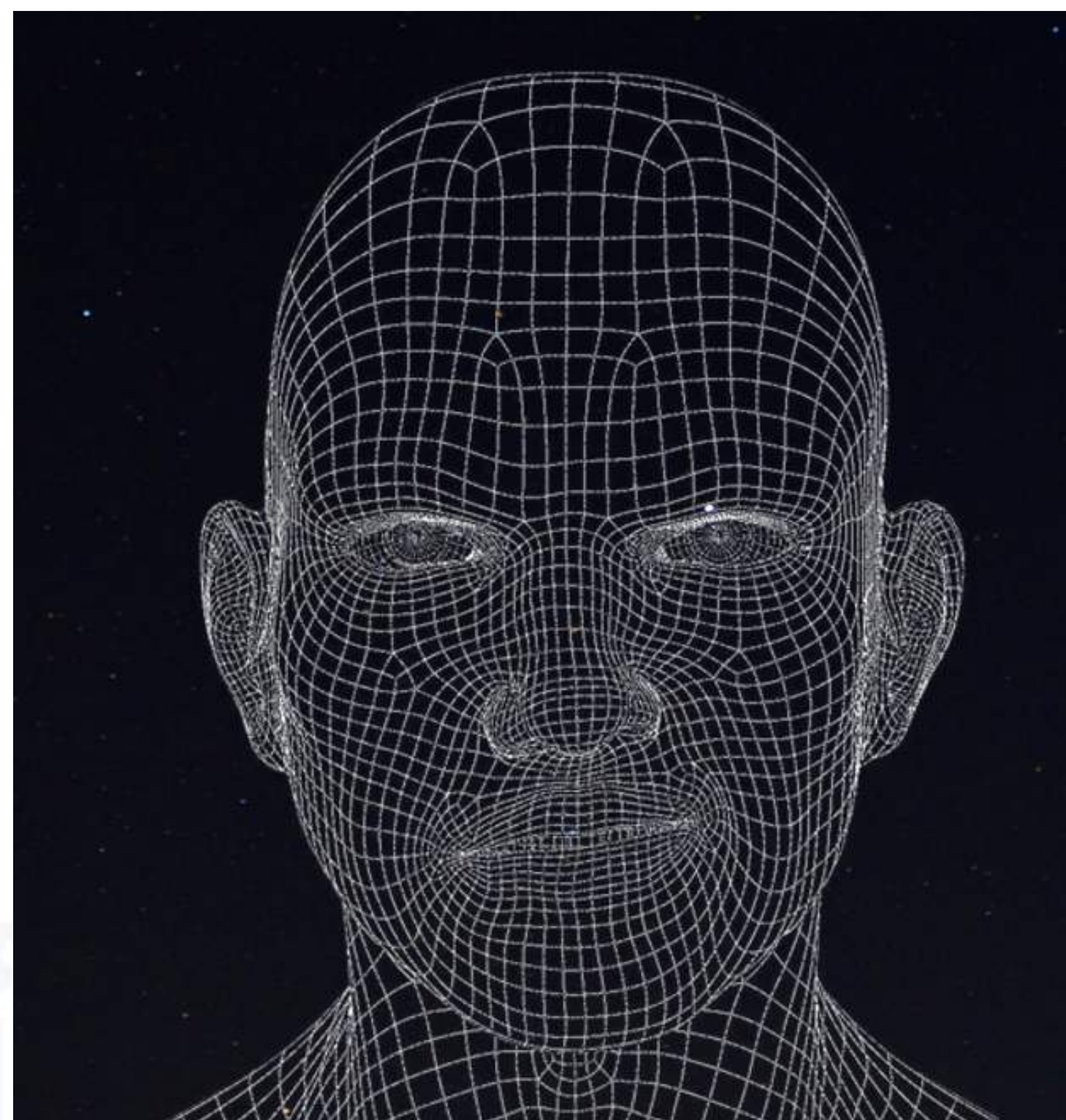


融合 CDN



API 加速

视觉-最重要的信息感知



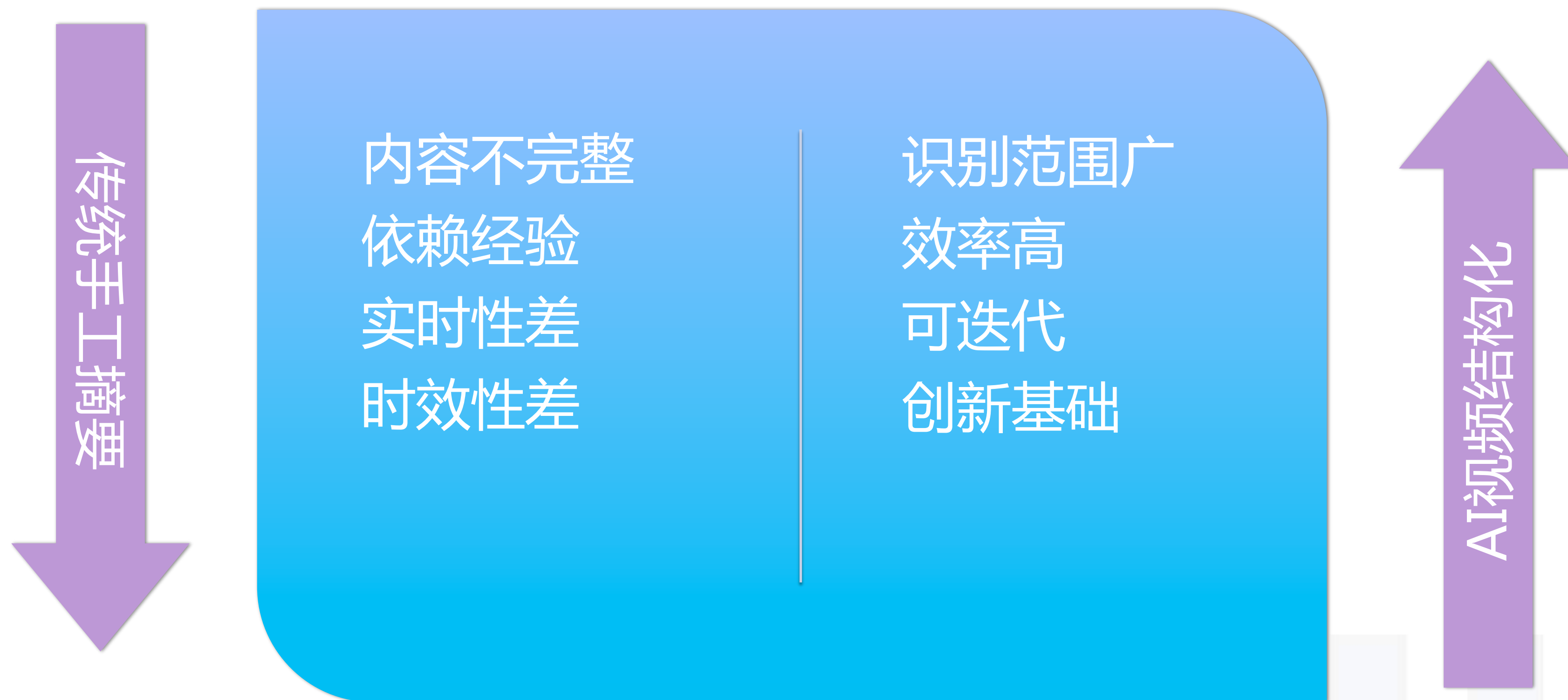
数据来源：根据2007年美国哈佛商学院的数据研究

2017中国网络视频用户情况



来源：《2017中国网络视听发展研究报告》

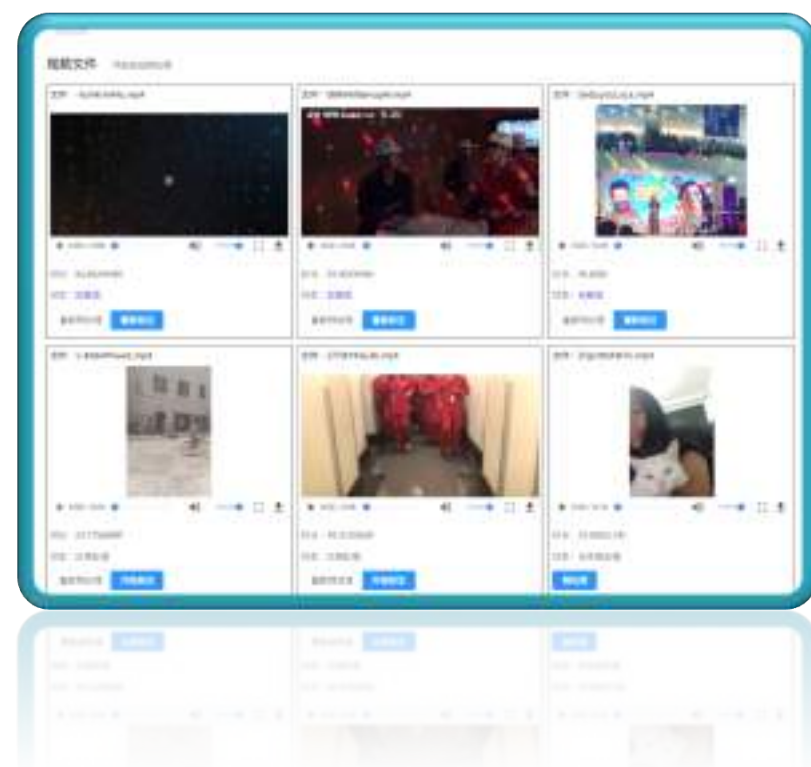
传统视频摘要 vs AI视频结构化



视频结构化场景



视频分解



视频截图 已标注

0:22 / 0:56

视频-ilcNK-6H4s.mp4

保存标签

视频标签

动作分类

视频片段标签

唱歌 烘焙 打篮球 开车

视频片段

唱歌		00:00:05.500		00:00:43.500	删除
烘焙		00:00:45.500		00:00:51.500	删除

基础模型要素



视频描述-标签



剪辑1： 01:02:03-01:10:05 体育新闻

事件1： 足球比赛

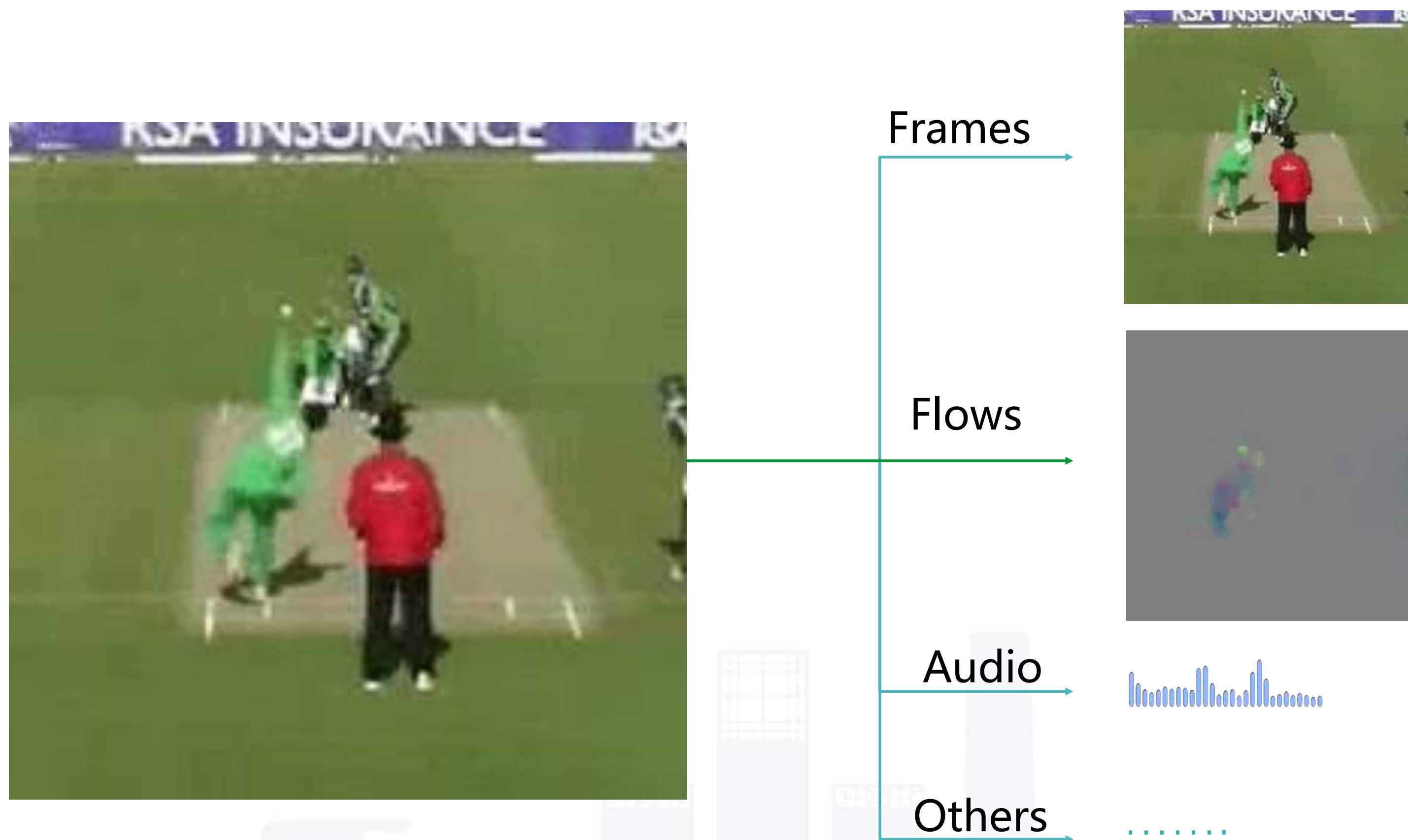
人： 梅西

物： 足球

球场(草坪)

文字： XX保险

视频处理

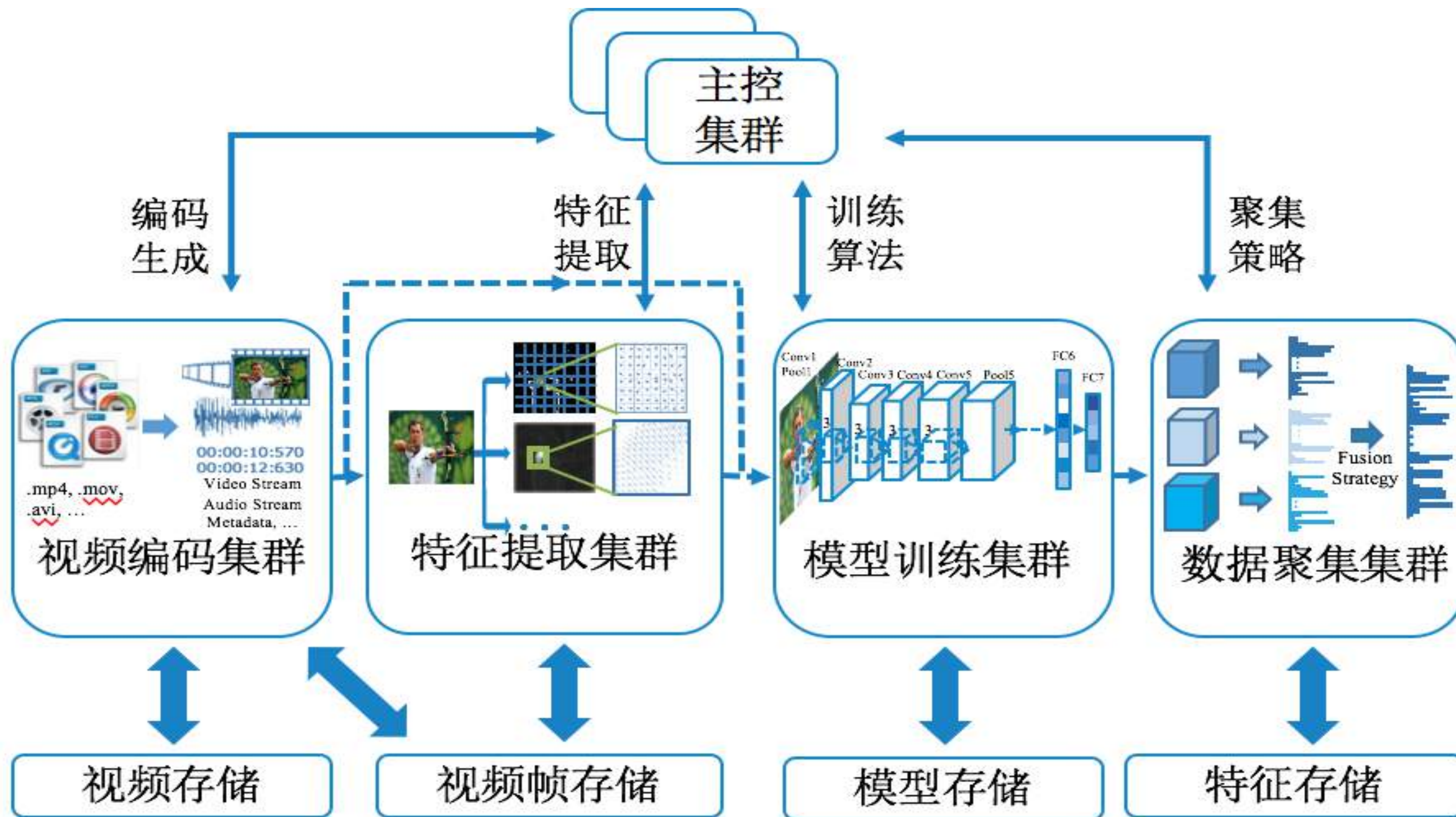


七牛云深度学习平台

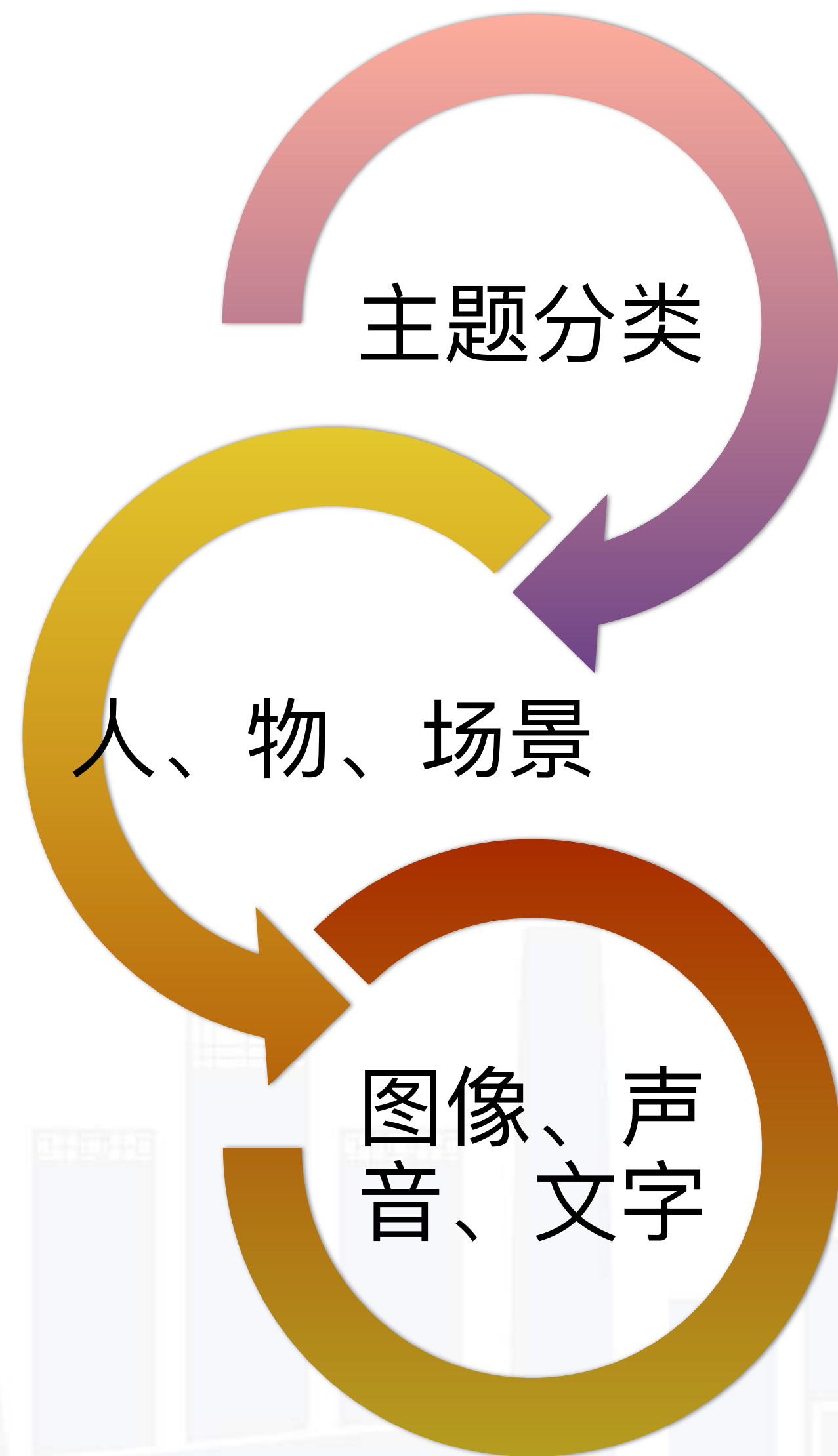
节省了80%的人力成本，缩短了50%的实现周期



大规模视频训练框架



结构化策略



主题分类-特征提取



ResNeXt

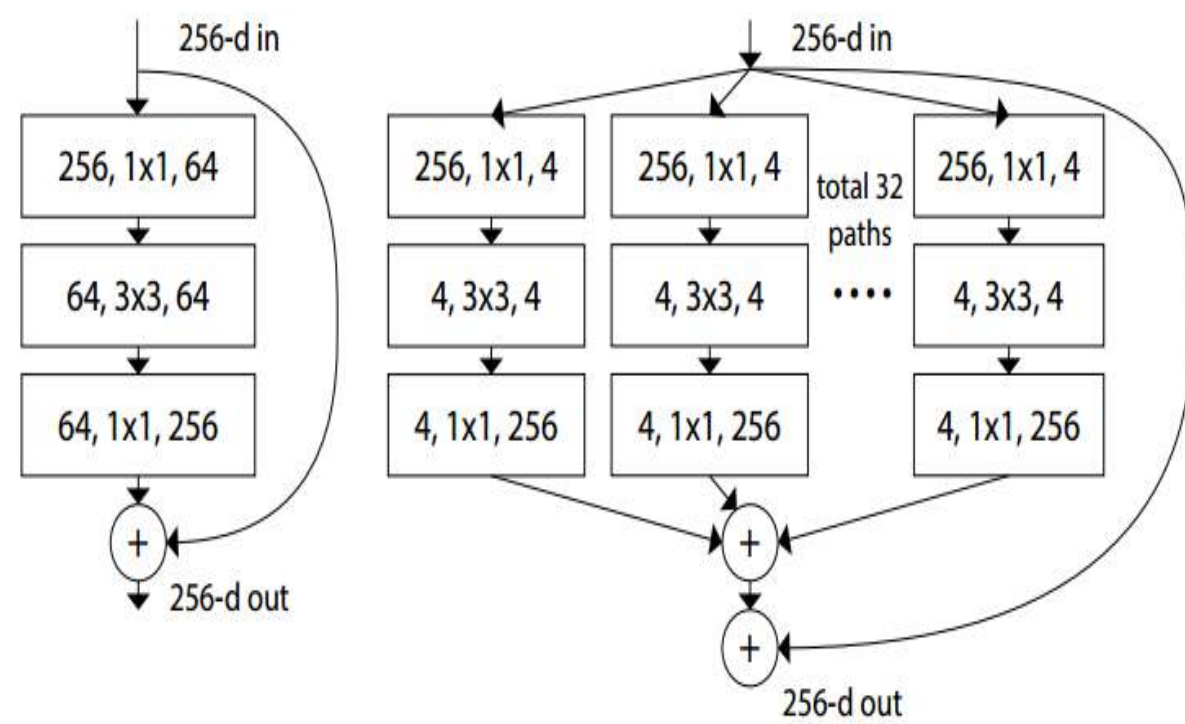
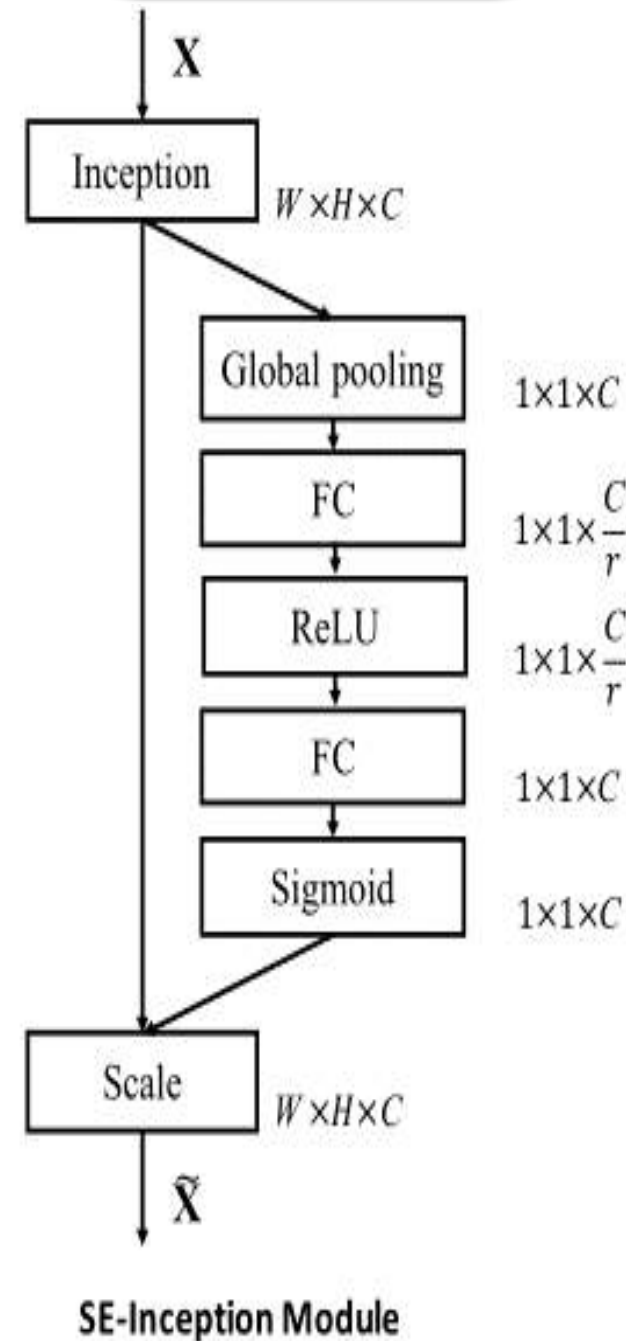
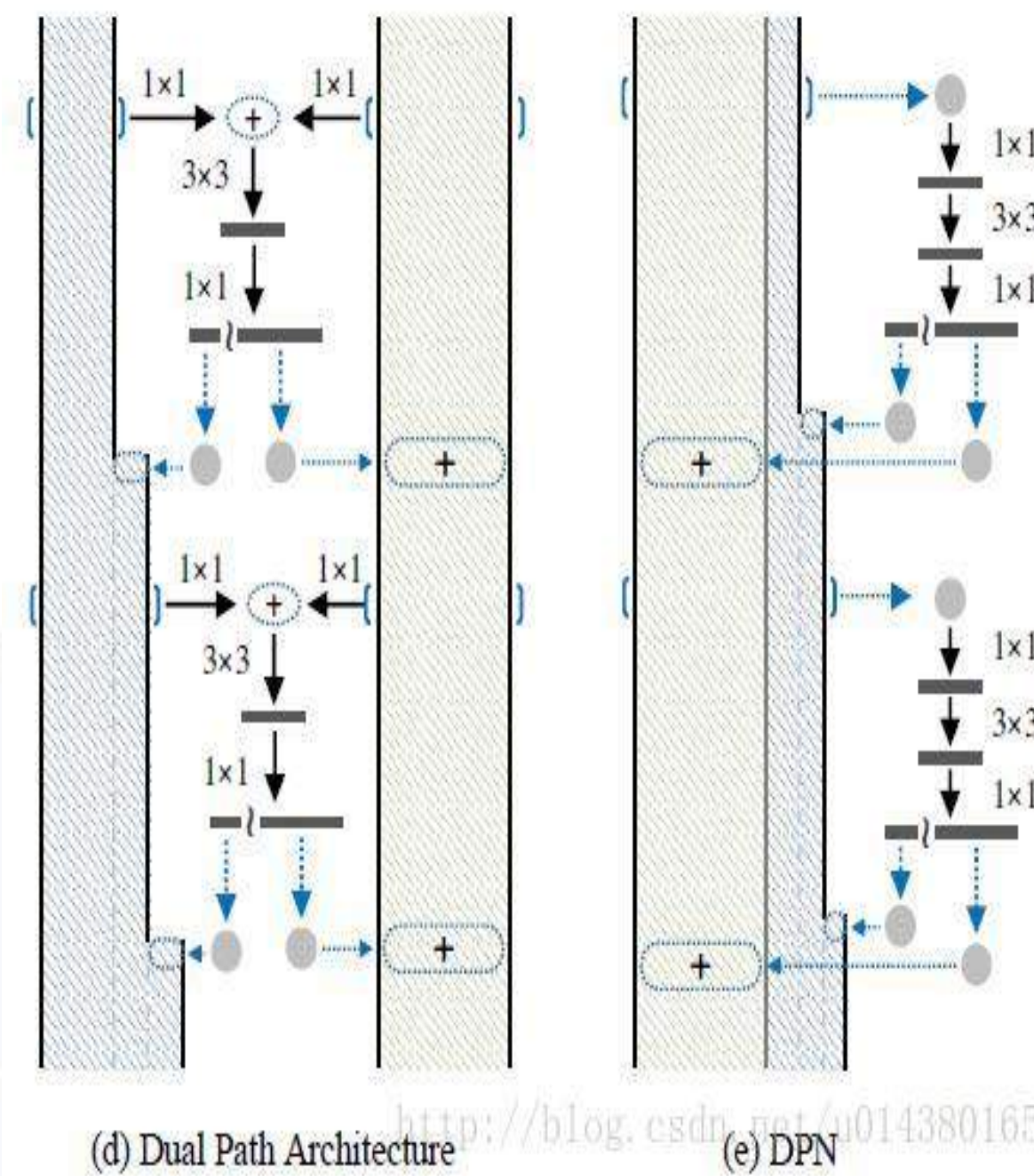


Figure 1. **Left:** A block of ResNet [13]. **Right:** A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

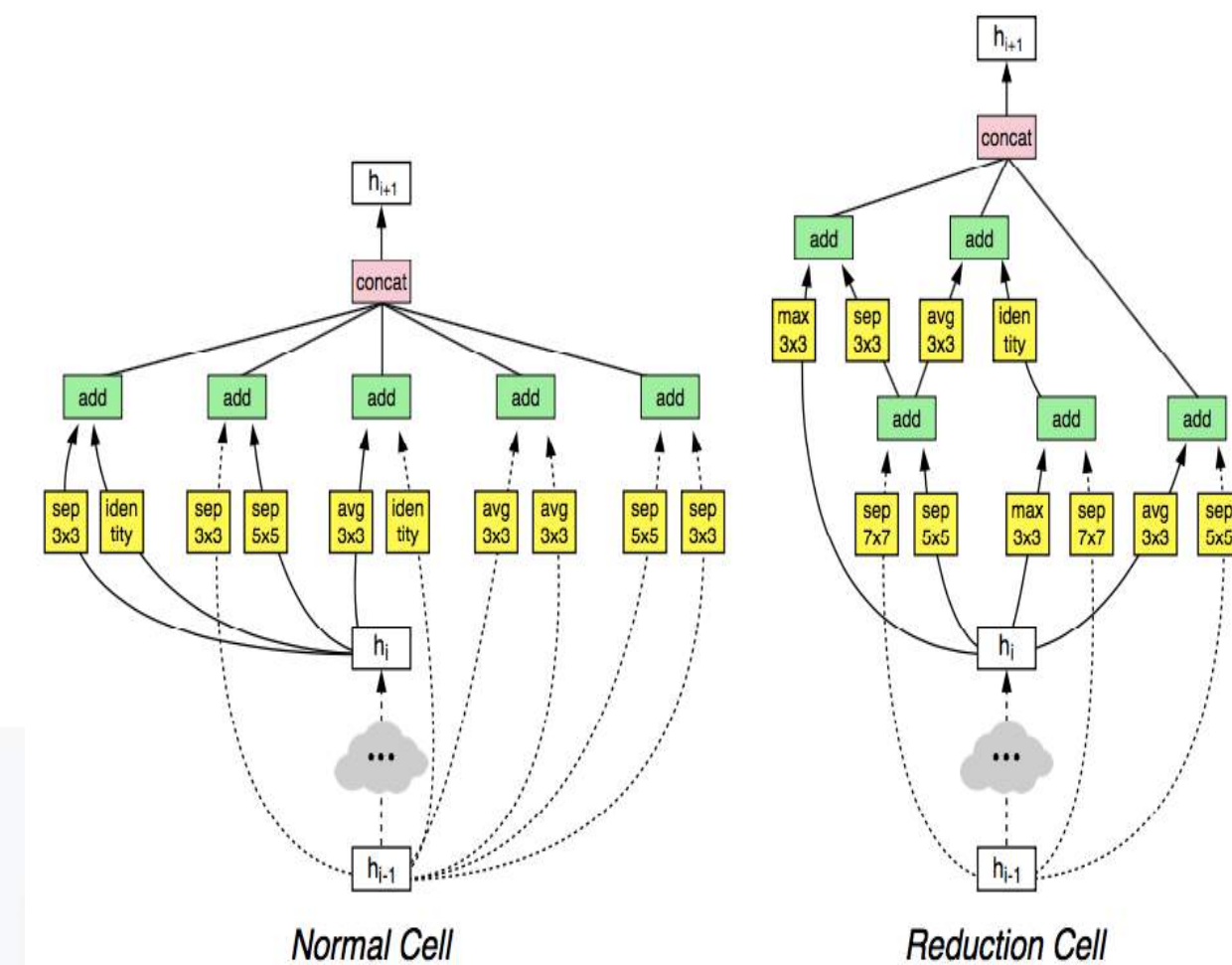
SENet



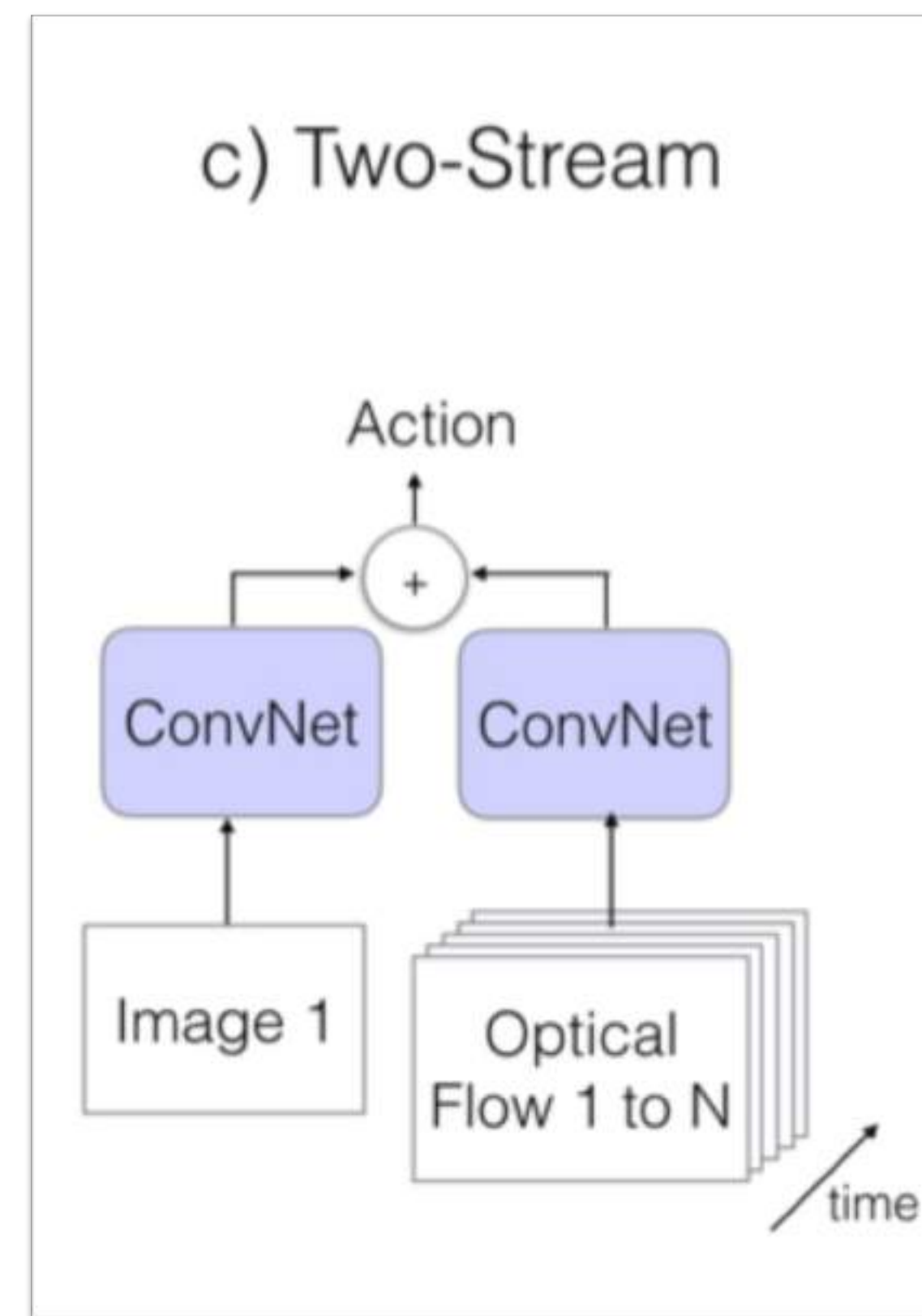
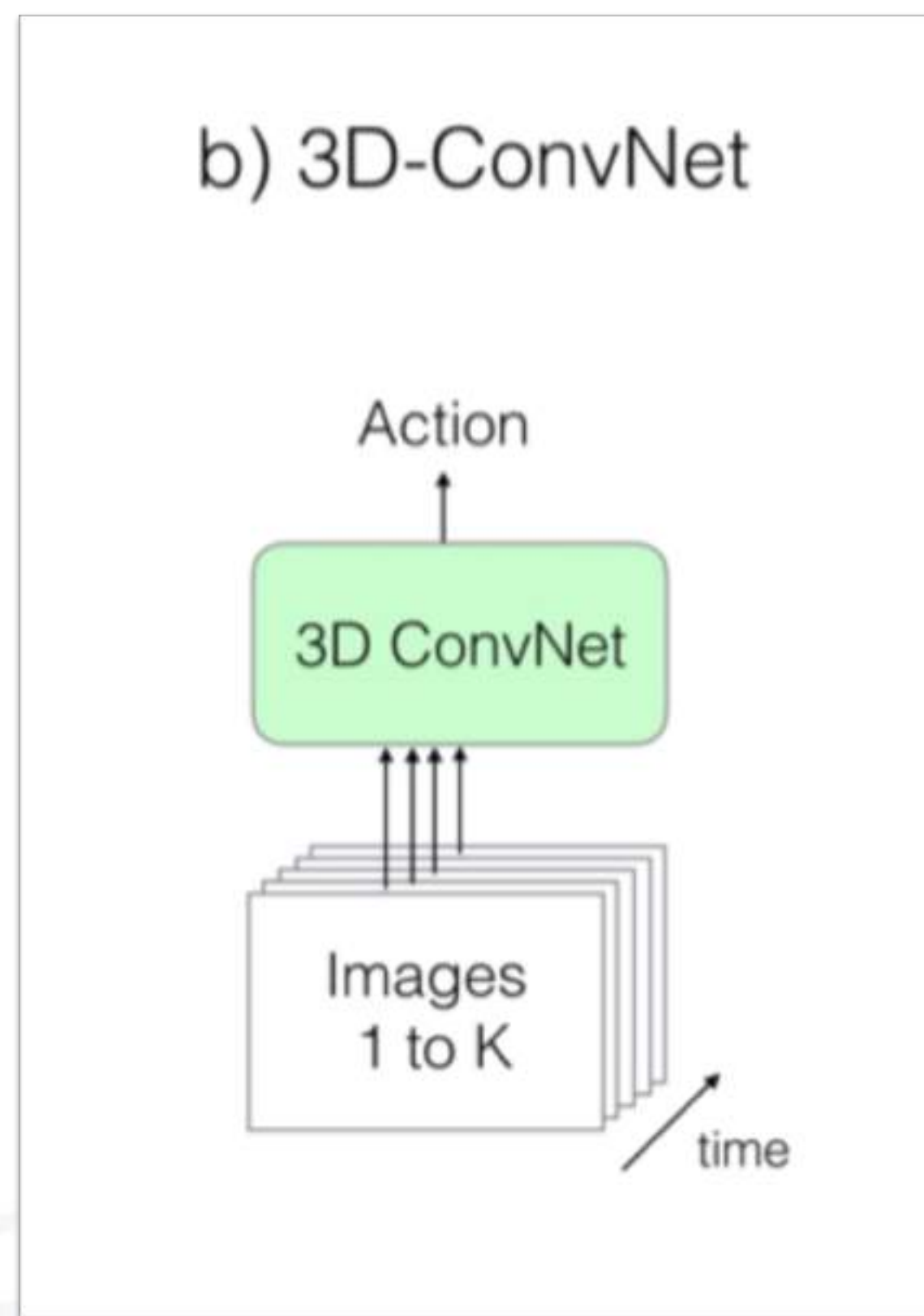
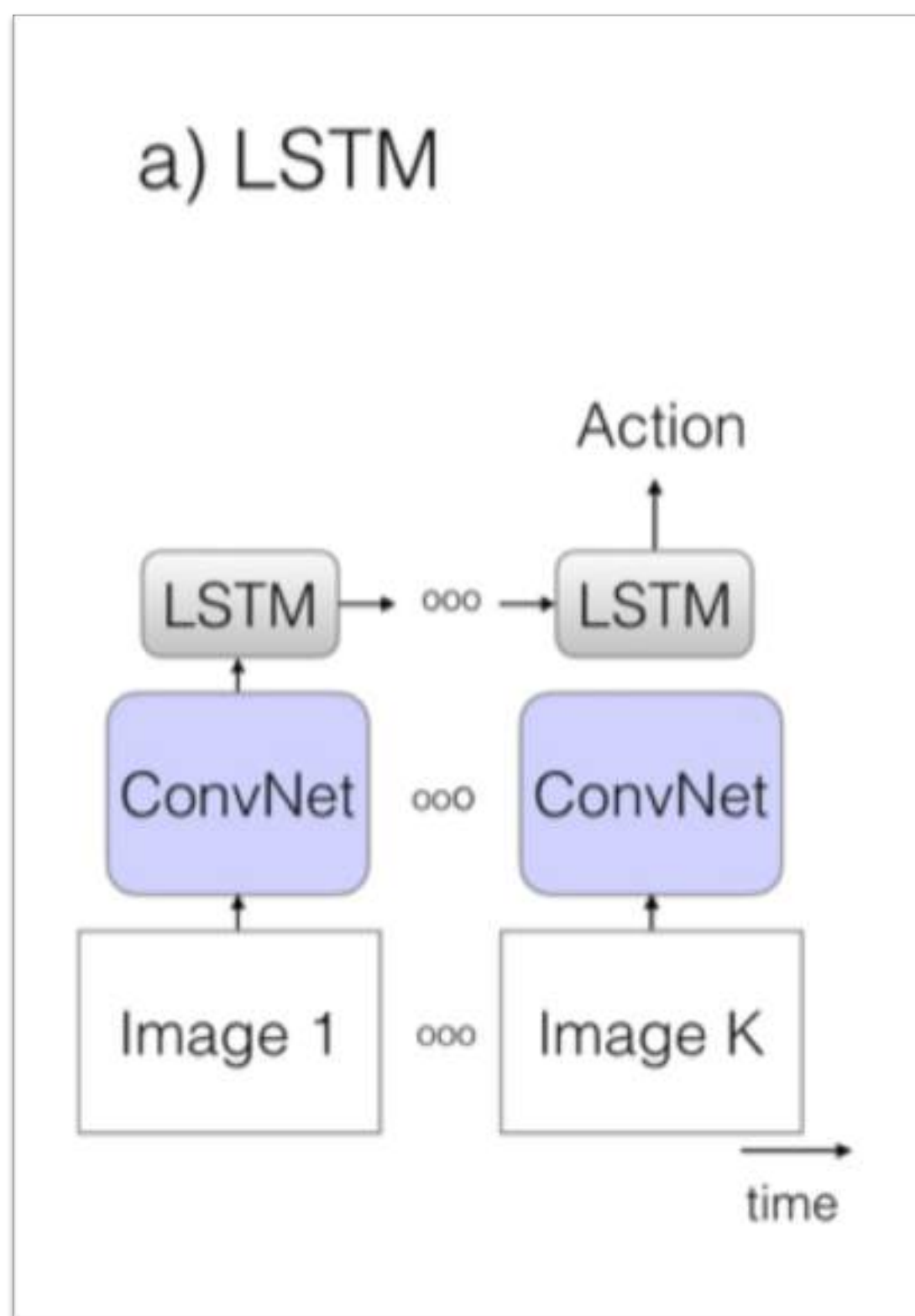
DPN



NASNet

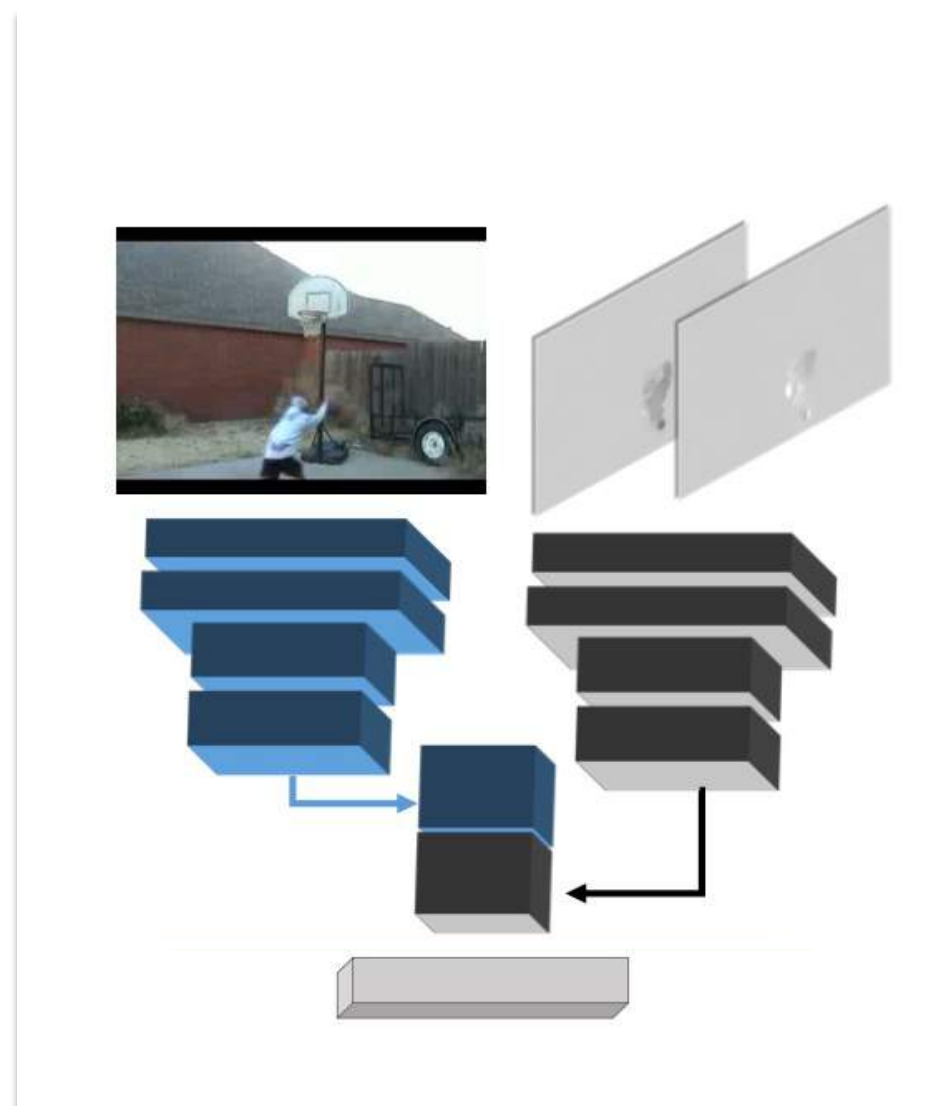


主题分类-模型训练



模型融合

a) Early fusion



b) Late fusion

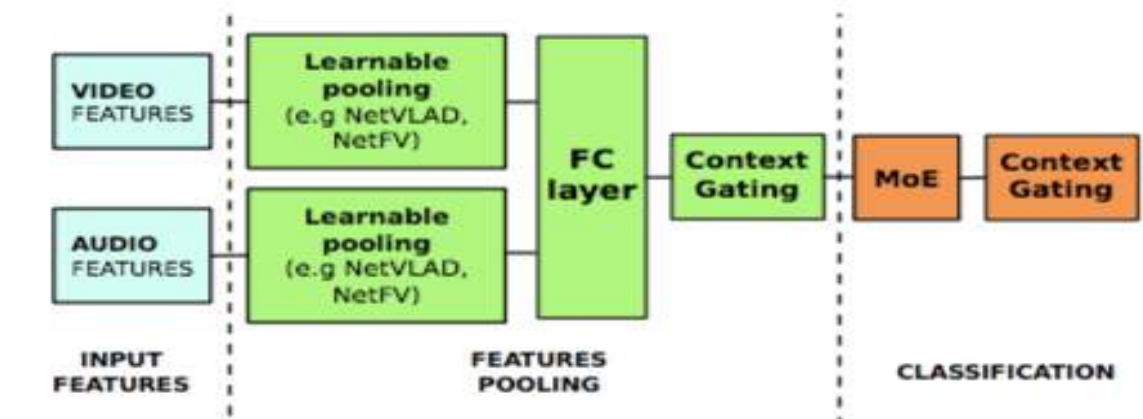
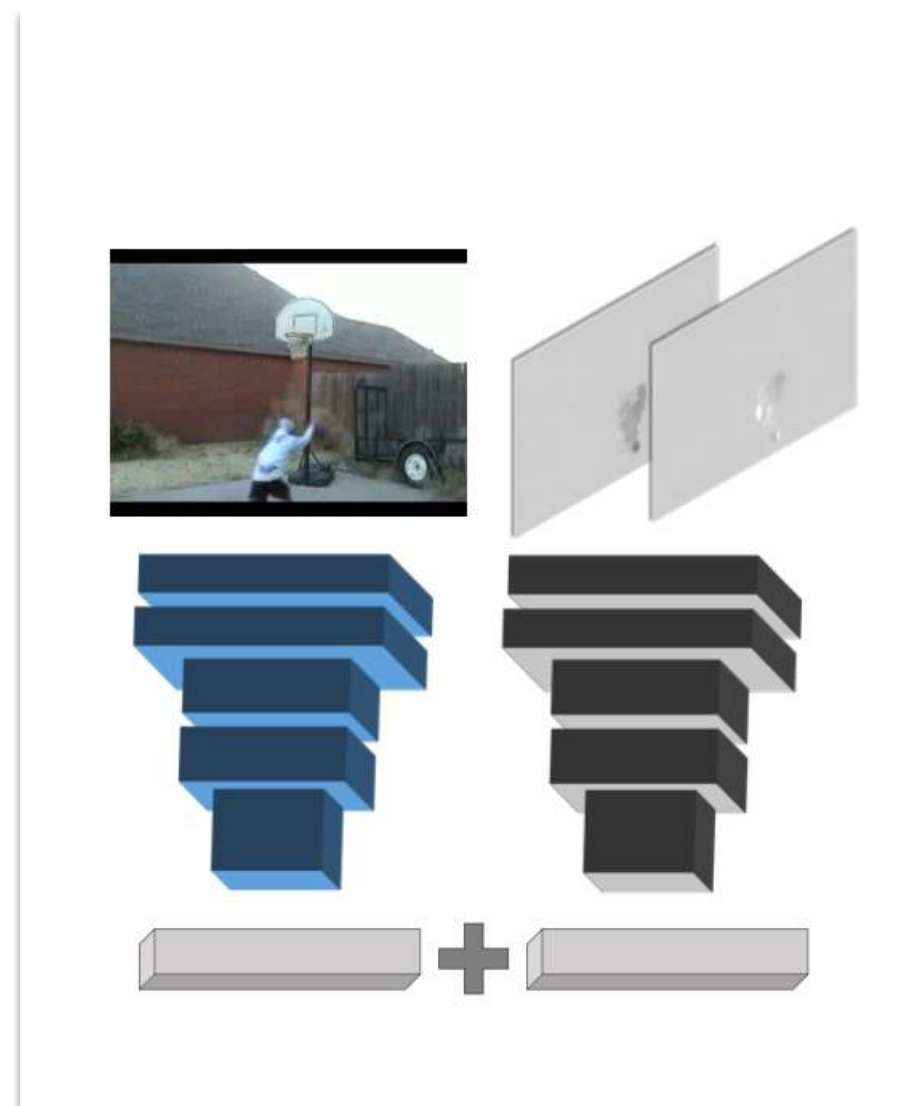


Figure 2: Overview of our network architecture for video classification (the “Late Concat” variant). FC denotes a Fully-Connected layer. MoE denotes the Mixture-of-Experts classifier [2])

$$Y = \sigma(WX + b) \circ X$$