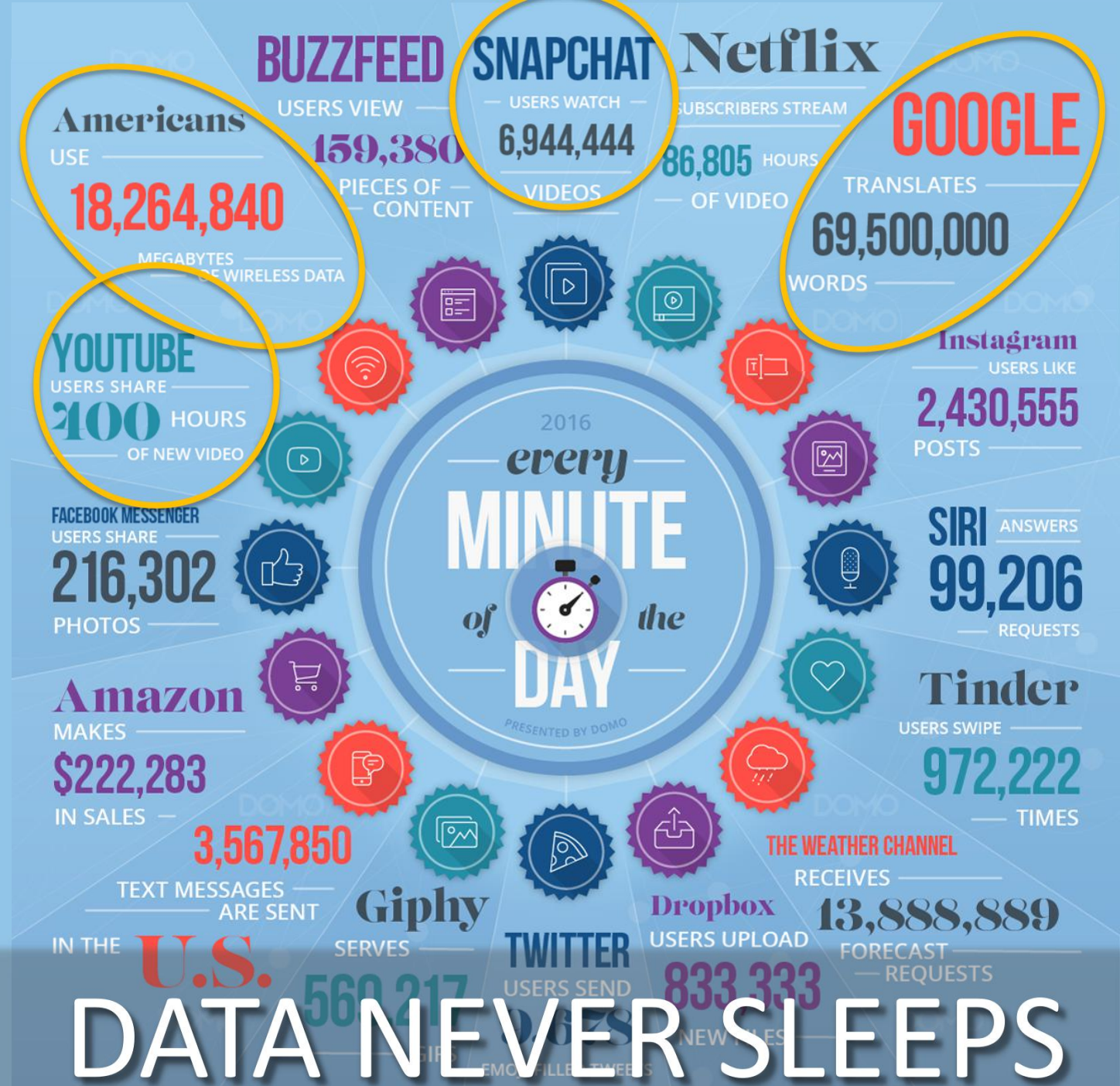




# What Makes a Video a Video: 视频理解

包怡欣

七牛云人工智能实验室



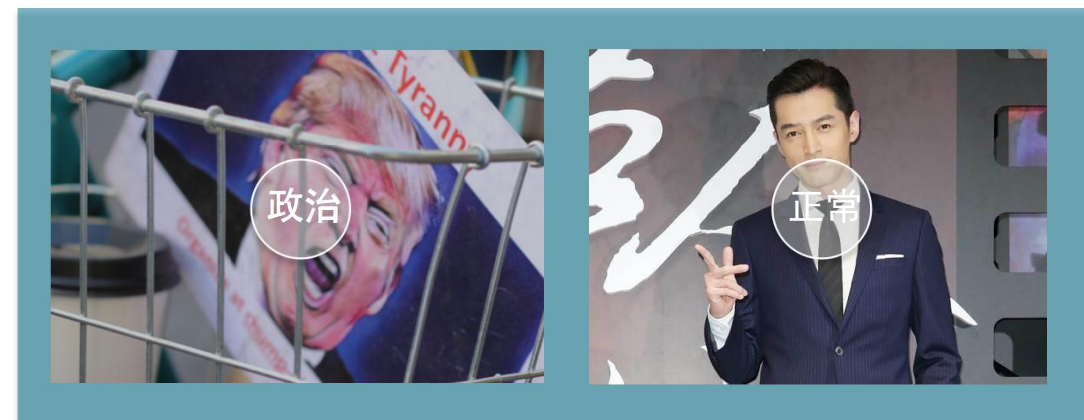
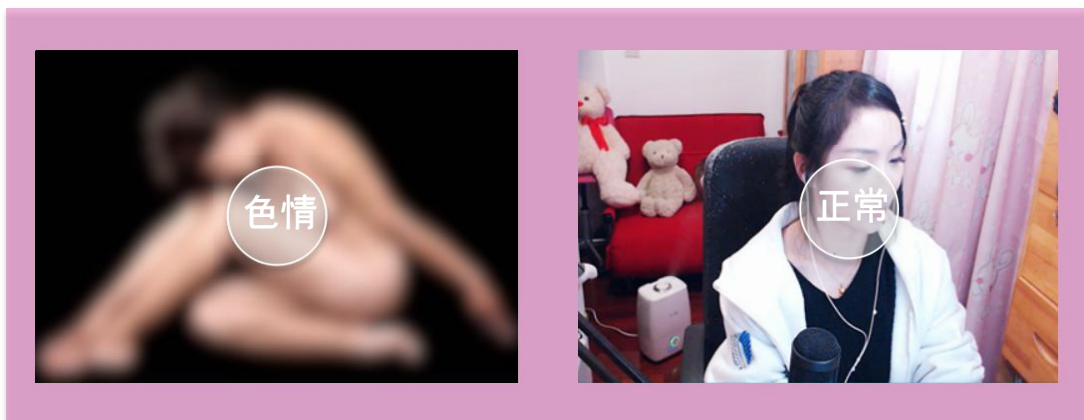
# DATA NEVER SLEEPS

# 视频社交





# 内容审核





↑ 视频列表 习近平

搜索: 习近平

视频文件	人物信息	人物出现频率	操作
 <p>(大国外交的苏州故事) 为世界和平护航 习近平讲述动人故事 [预编版].mp4</p>	 <p>习近平</p>		<a href="#">详情</a>
 <p>习近平：不要人夸颜色好 只留清气满乾坤 [字幕版].mp4</p>	 <p>习近平</p>		<a href="#">详情</a>
 <p>习近平出席金砖国家工商论坛开幕式并发表主旨演讲[预编版].mp4</p>	 <p>习近平</p>		<a href="#">详情</a>
 <p>demo.mp4</p>	 <p>习近平</p>		<a href="#">详情</a>

# 视频检索

# 视频检索



七牛视频结构化

视频列表

首页 / 视频详情


视频详情

视频列表

请输入你要搜索的人物

详情

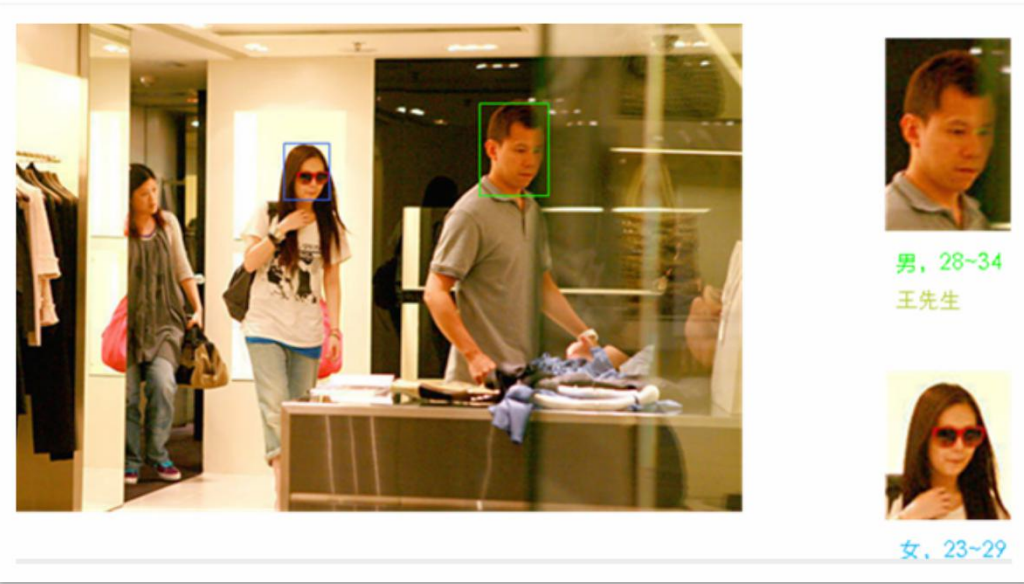
普通人



视频结构化分析结果展示：左侧为视频缩略图，右侧为时间轴，时间轴上显示了视频中出现的普通人的面部特征识别结果，用绿色条块表示。



# 中国的摄像头数量



# 视频广告植入





# 视频广告植入



# 自动驾驶



# 体育赛事分析





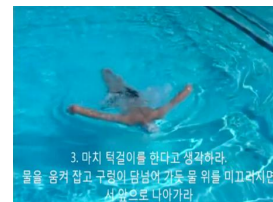
# 运动相关



健身



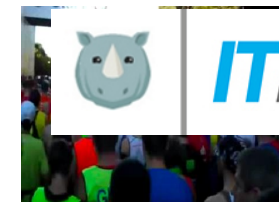
篮球



游泳



滑板



马拉松



羽毛球



乒乓球



排球



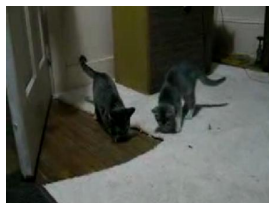
足球



高尔夫

## VIDEOS ARE EVERYWHERE

# 动物相关



猫



狗



牛



鸟



海豚



熊猫



兔子



仓鼠



长颈鹿



猿猴

# 食物相关



沙拉



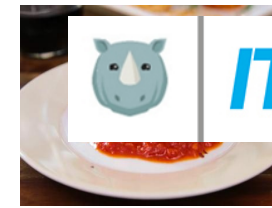
面包



饺子



生煎



意面



咖啡



Pizza



冰淇淋



蛋饼



牛排

## VIDEOS ARE EVERYWHERE

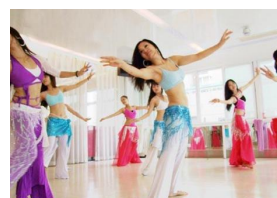
# 艺术相关



吉他



芭蕾舞



肚皮舞



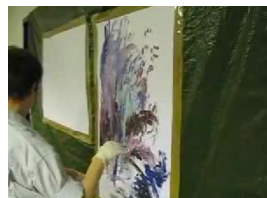
大提琴演奏



合唱队



歌剧



绘画



钢琴



爵士舞



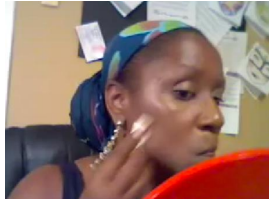
演唱会



# 美妆相关



面部按摩



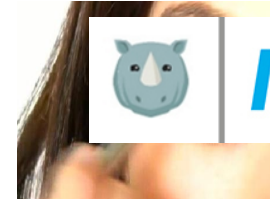
卸妆



卷发



抹口红



画眼妆



面膜



做指甲



底妆



刮胡子



洗牙

## VIDEOS ARE EVERYWHERE

# 休闲相关



晒娃



野营



旅行



钓鱼



放风筝



野餐



玩水枪



唱K



阅读



广场舞





# What Makes a Video a Video

# 视频的时序关联性

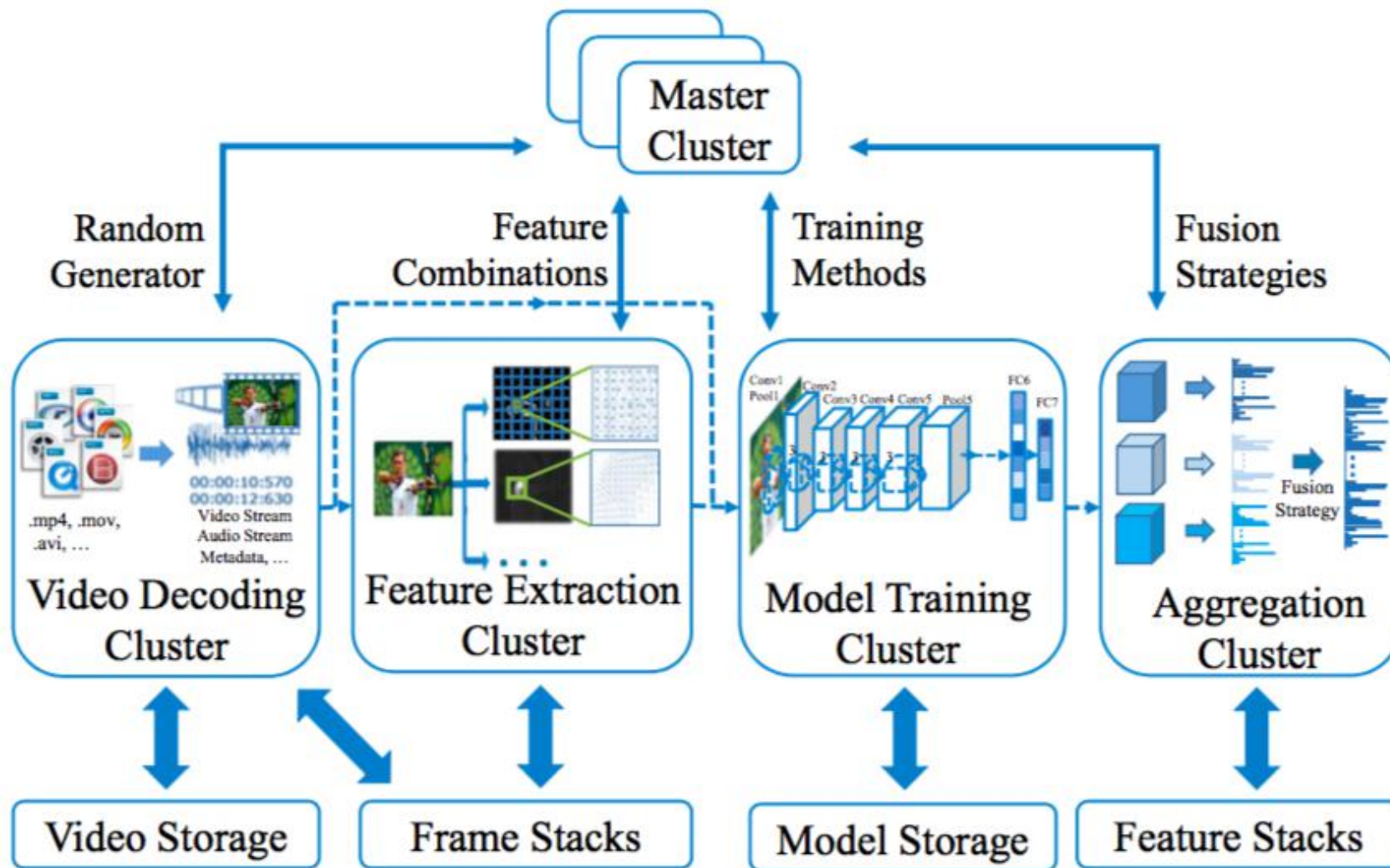


# 视频的阶段性





# 大规模视频训练平台



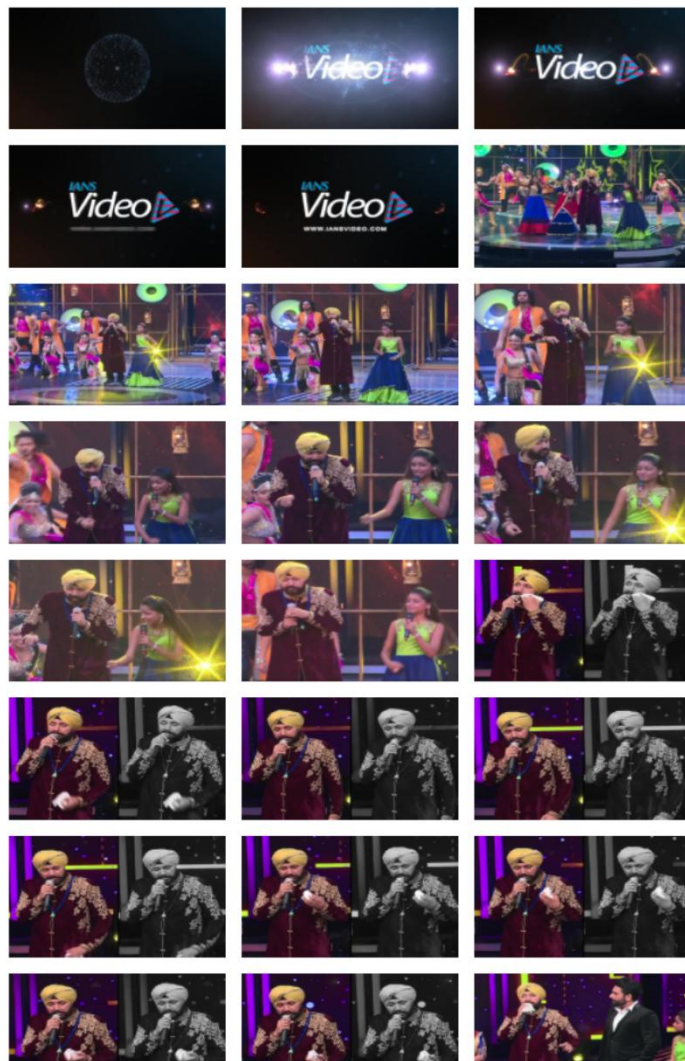
# 视频标注系统



视频文件 开始自动预处理

文件: -ilcNK-6H4s.mp4 时长: 56.842444秒 状态: 已标注 重新预处理 重新标注	文件: 0BRWf5aHupM.mp4 时长: 54.404344秒 状态: 已标注 重新预处理 重新标注	文件: 0wQzytULxLk.mp4 时长: 49.83秒 状态: 已标注 重新预处理 重新标注
文件: 1-XbbiPPwWE.mp4 时长: 24.775689秒 状态: 已预处理 重新预处理 开始标注	文件: 277kYK5u3E.mp4 时长: 49.412056秒 状态: 尚未预处理 重新预处理 开始标注	文件: 2FgIttXW7U.mp4 时长: 19.690511秒 状态: 尚未预处理 重新预处理 开始标注

视频帧 已标注



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

# 数据处理



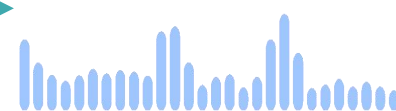
Frames



Flows



Audio



Others

.....



# 特征提取



## ResNeXt

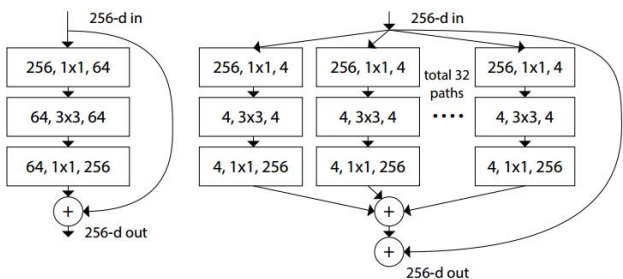
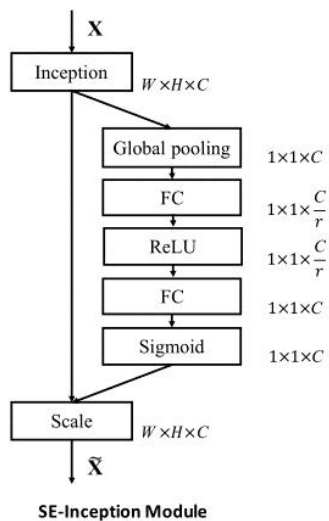
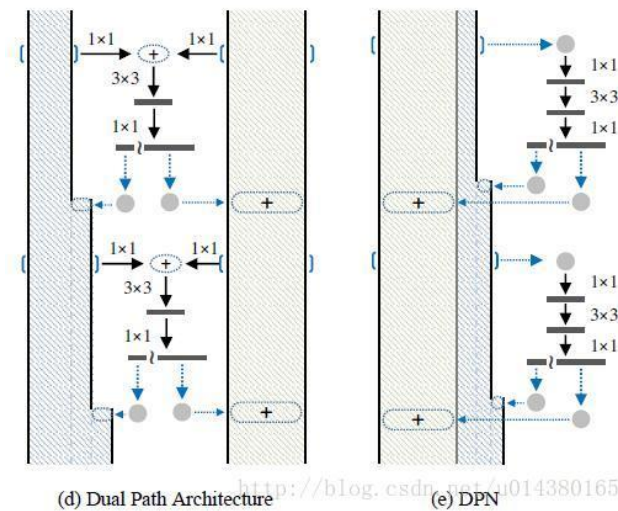


Figure 1. **Left:** A block of ResNet [13]. **Right:** A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

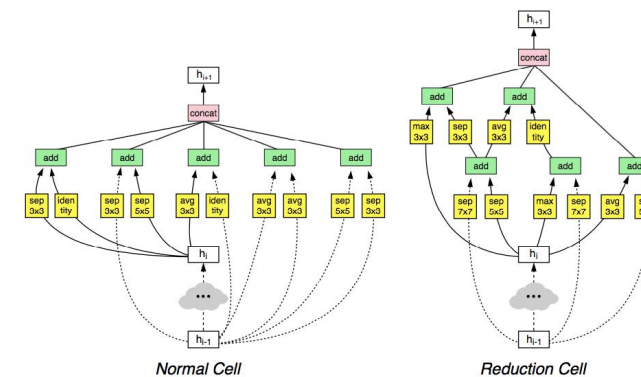
## SENet



## DPN



## NASNet

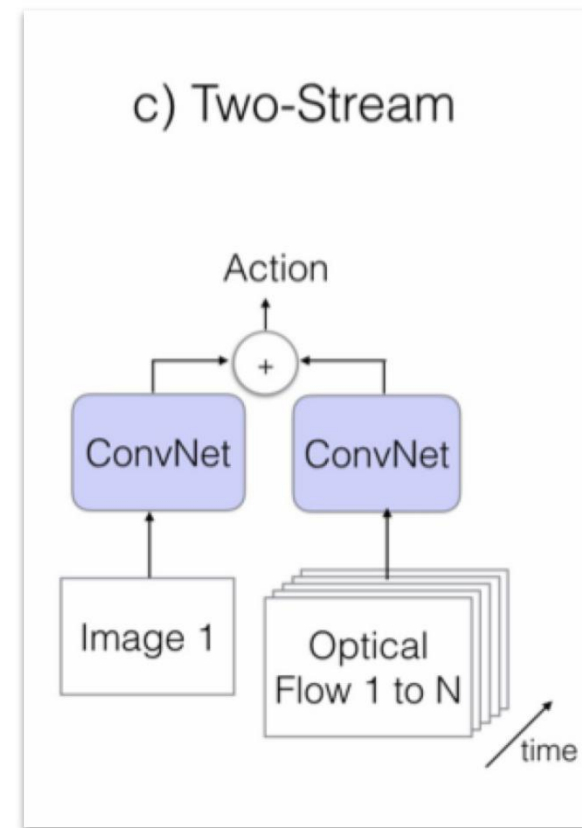
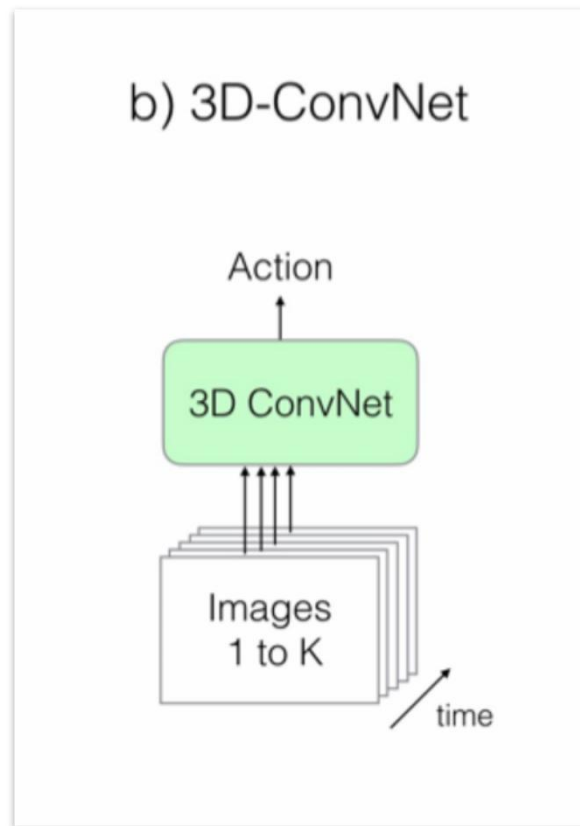
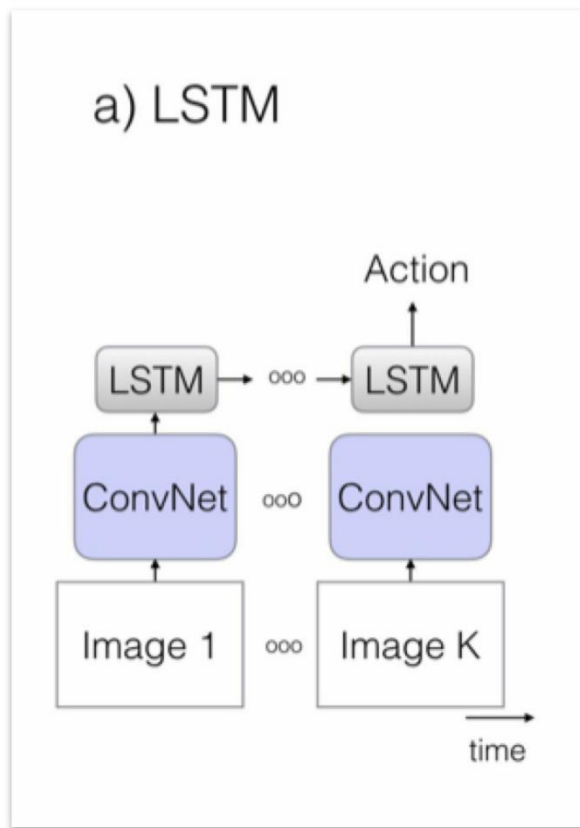




# 特征提取

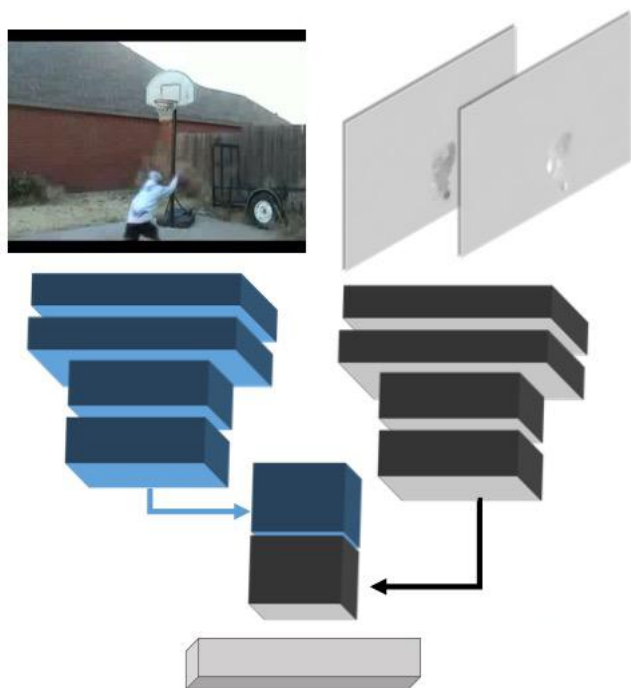


# 模型训练

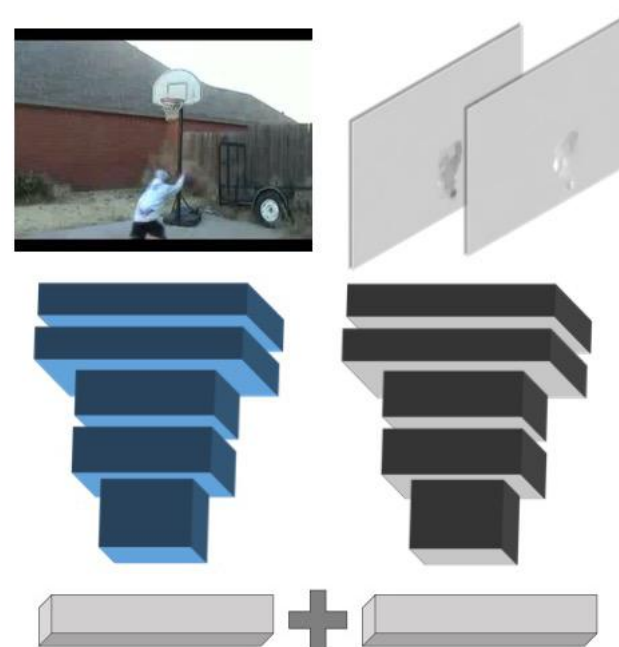




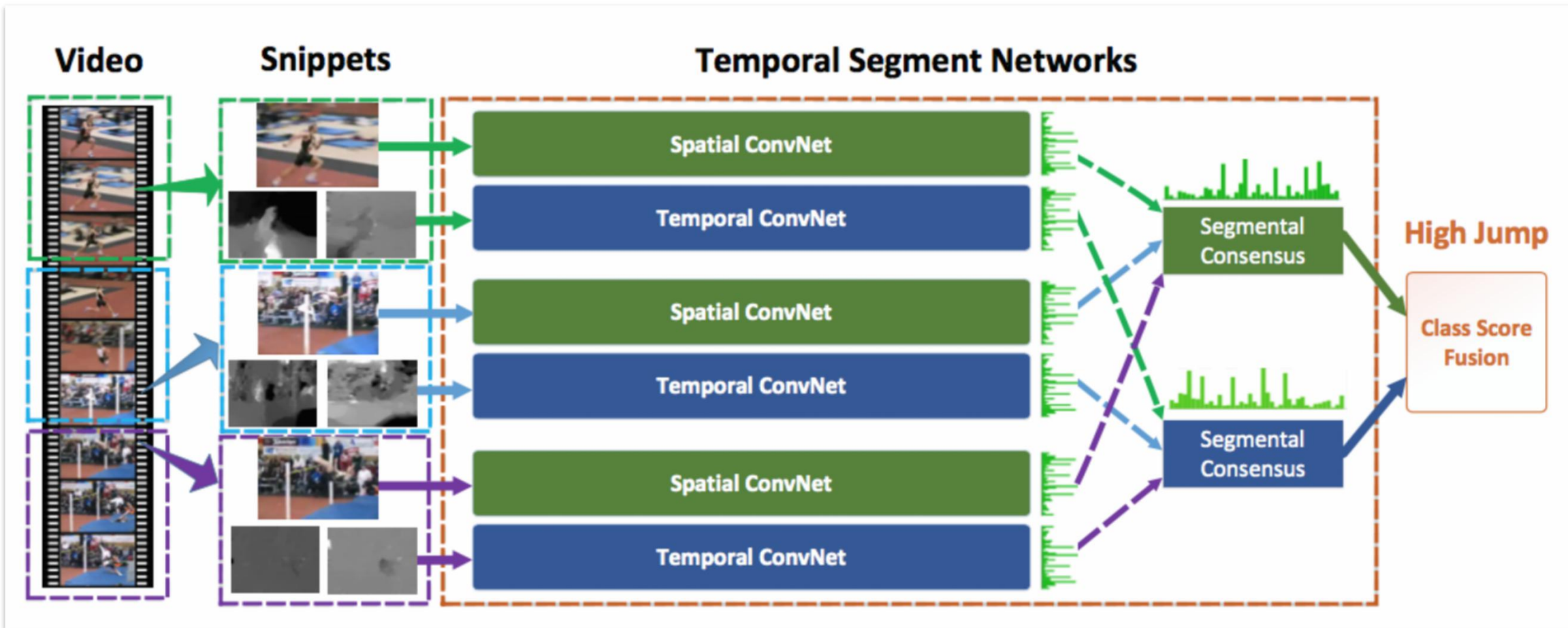
a) Early fusion



b) Late fusion



# 举个例子



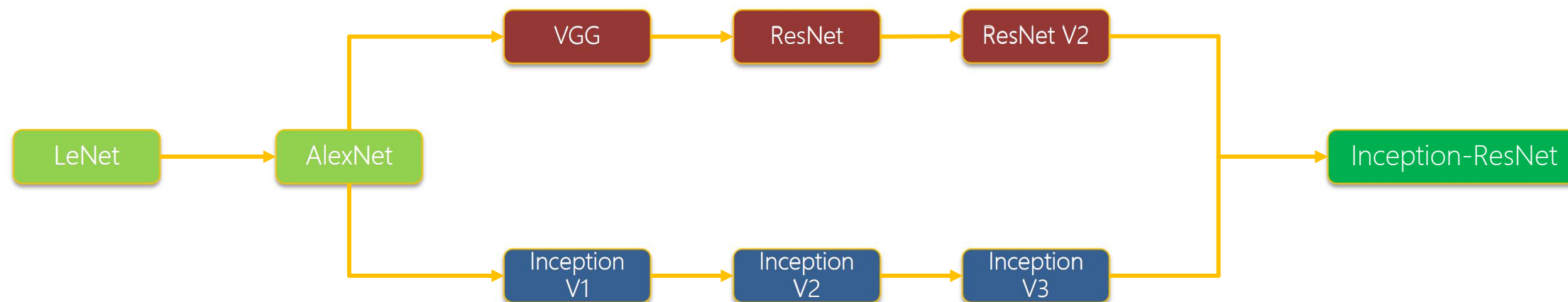






# 关于训练

- 保持对数据的敬畏之心
- 模型多样性的重要性
- 研究模型的结构，直击要害





# Thanks!

## Q & A