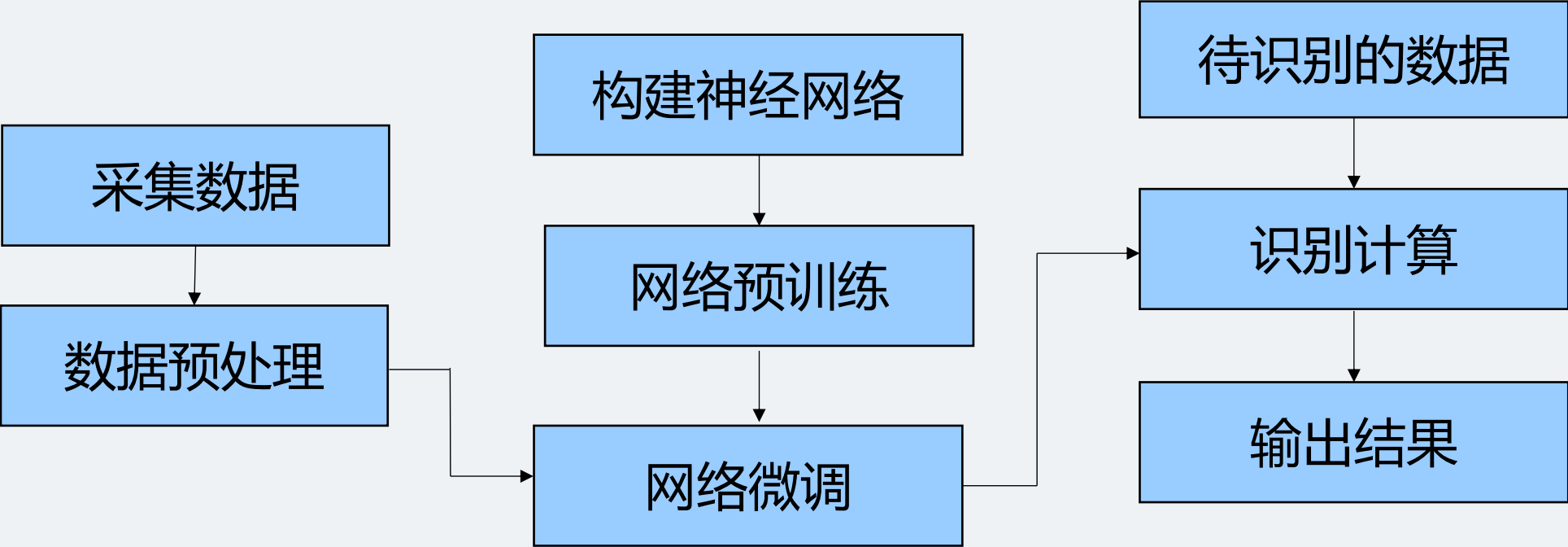
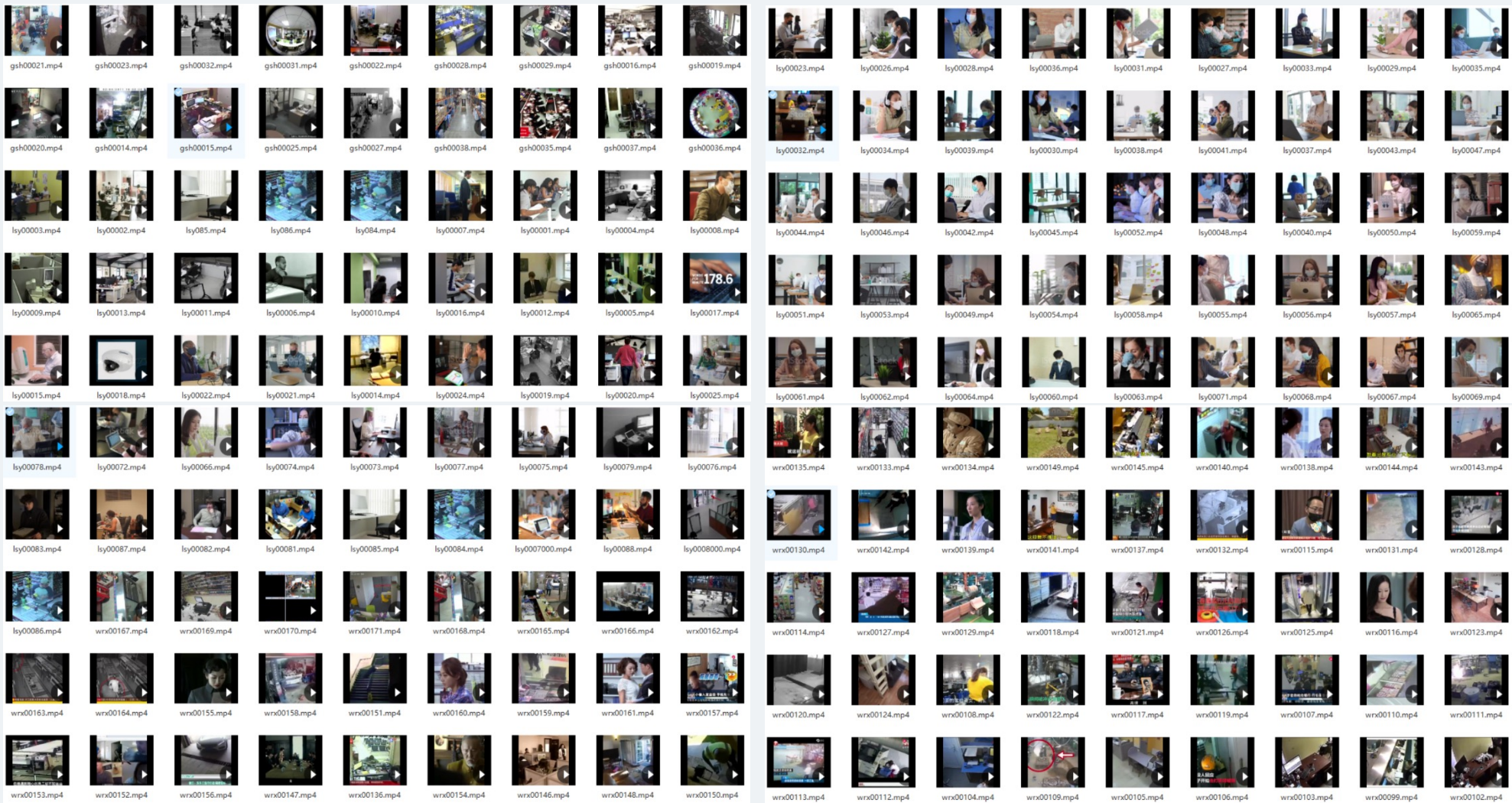




# 项目进展情况

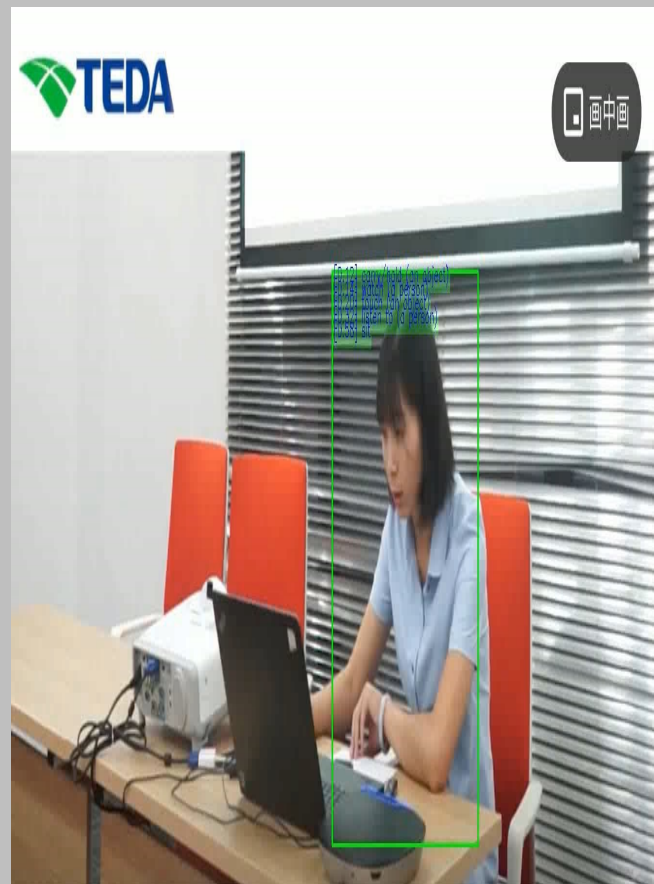




不同复杂场景下的行人离岗视频

# 项目进展情况

## 基于深度学习的复杂状态离岗检测



上面三个demo为经过我们的网络模型处理后的视频，识别框左上角给出了可能的每个动作的概率



# 项目实施问题及解决措施

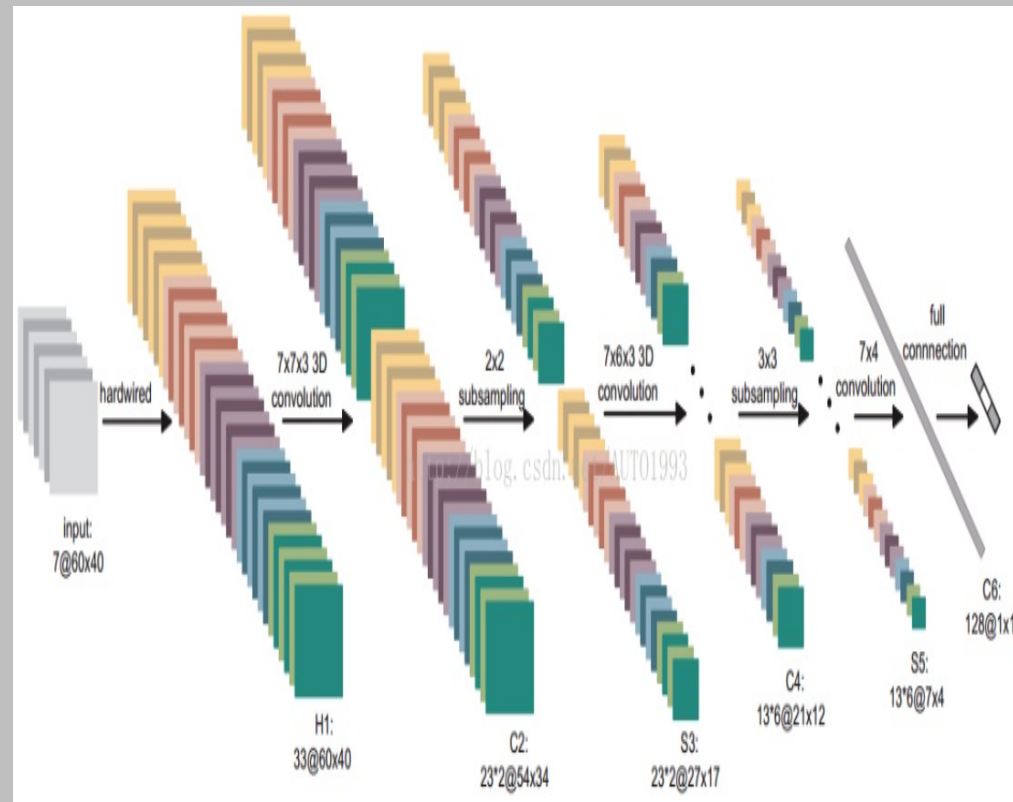


# 问题一

## 3D-CNN



单帧图像未结合时序信息易发生误判的情况

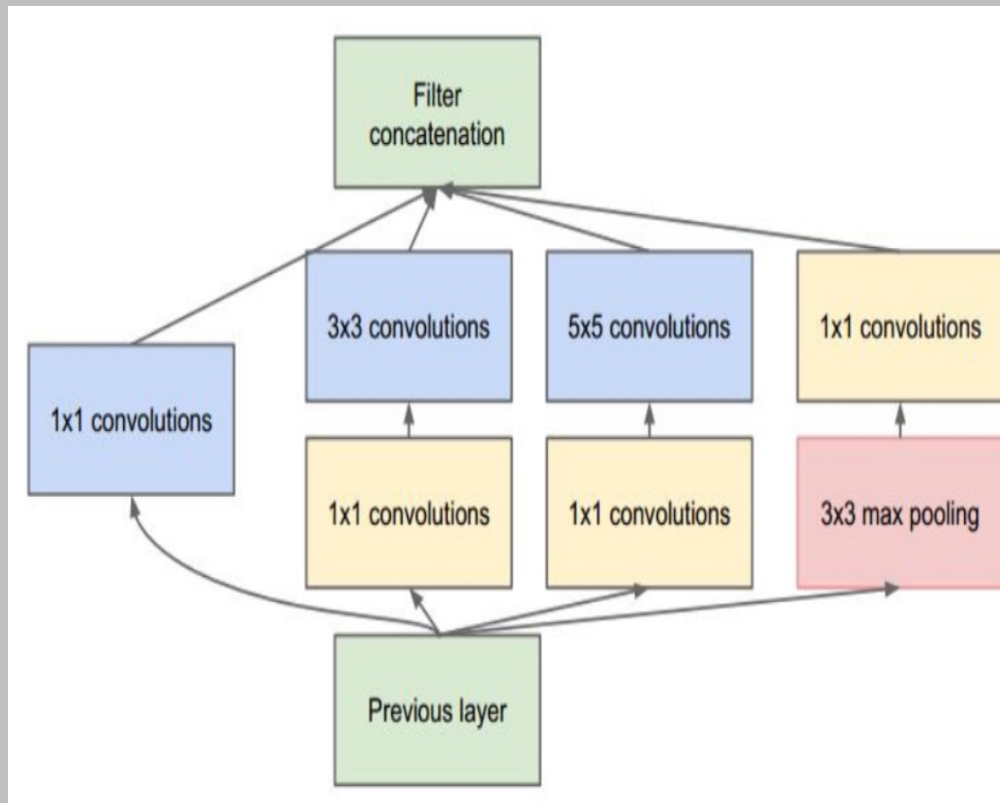


## 问题二



动作识别只关注视频中的人的特征，忽略了人与背景中某些属性之间的关系

## 2D-CNN



## 基于深度学习的复杂状态离岗检测

# 消融学习



# 如何快速的提取特征，提高计算效率，节约计算资源，实现实时的岗位监测

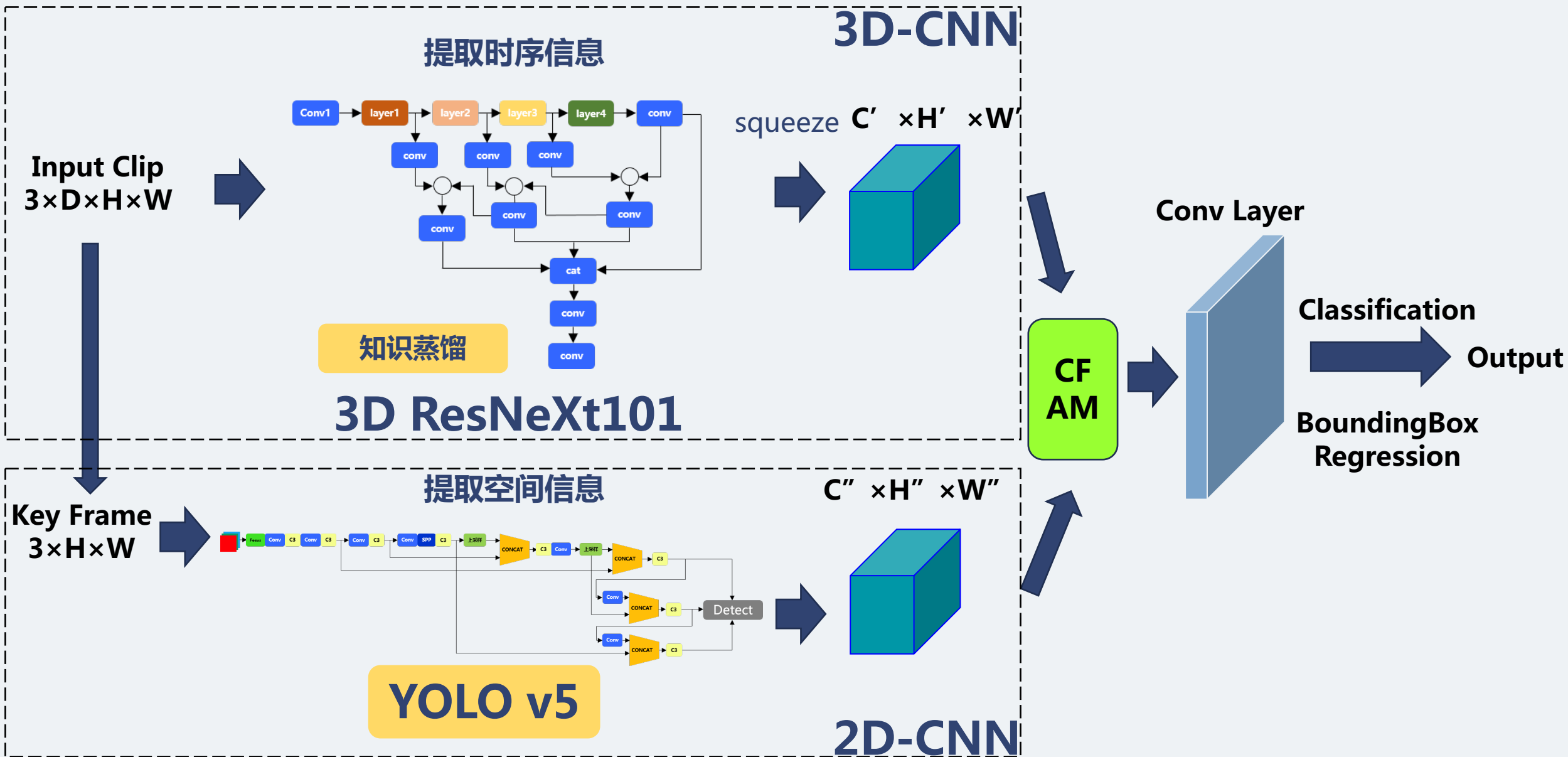
	Model	UCF101-24	J-HMDB-21
	2D	61.7	36.0
	3D	71.5	41.5
	2D + 3D	77.9	47.1
	2D + 3D + CFAM	<b>85.8</b>	<b>64.9</b>

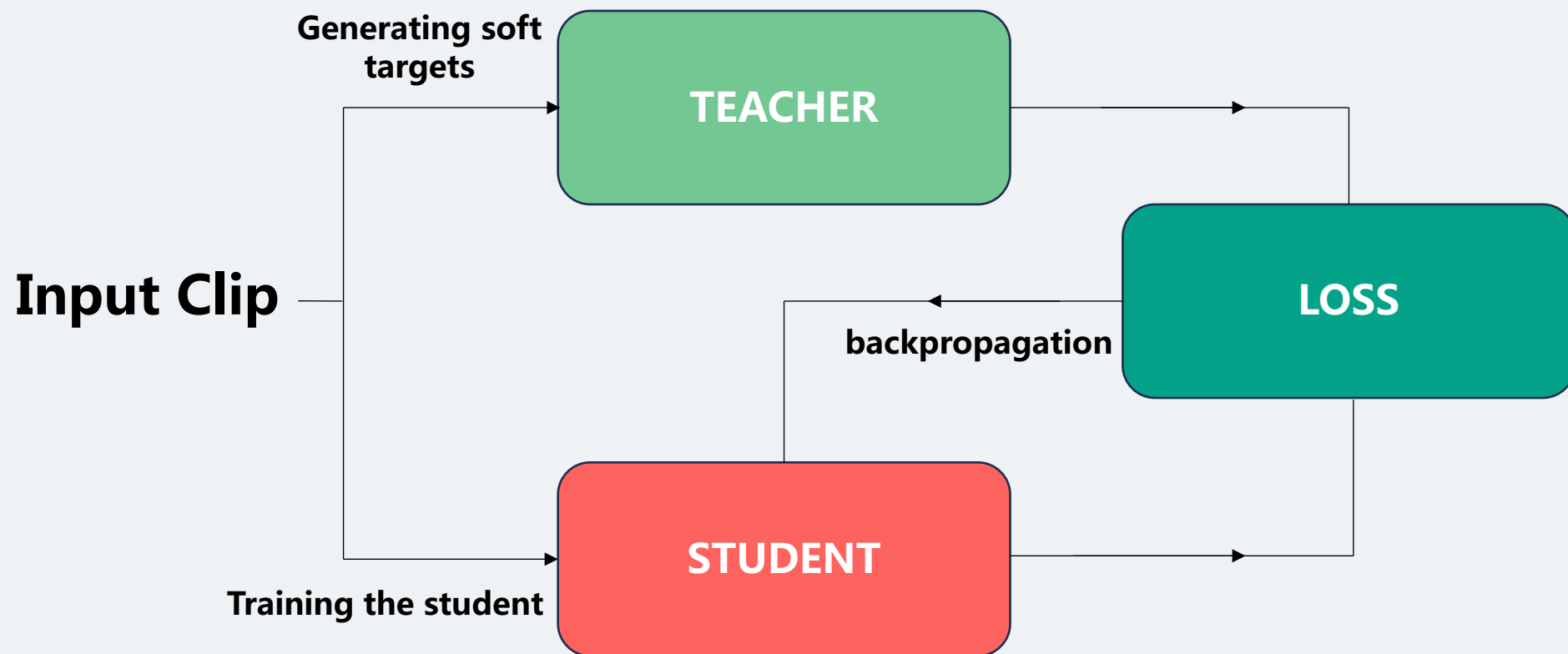
  

	Model	Localization (recall)	Classif.
UCF101-24	2D	91.7	85.9
	3D	90.8	92.9
	2D + 3D	93.2	93.7
	2D + 3D + CFAM	<b>93.5</b>	<b>94.5</b>
J-HMDB-21	2D	94.3	50.6
	3D	76.3	69.3
	2D + 3D	94.5	63.0
	2D + 3D + CFAM	<b>97.3</b>	<b>76.1</b>

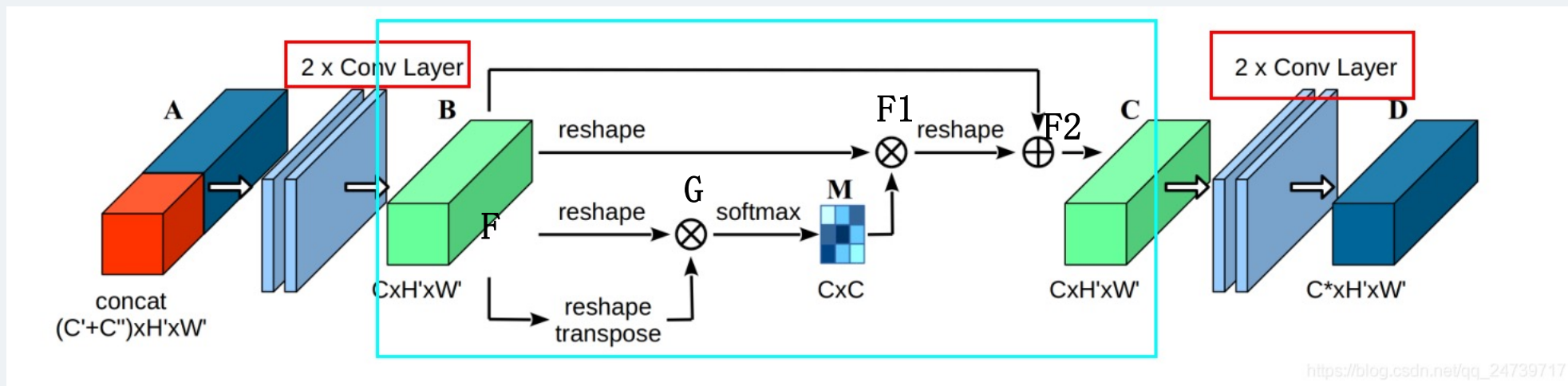


# 基于YOWO的2D+3D混合模型





使用知识蒸馏方法来训练3D CNN，从而使网络能够学习长距离的类似光流的运动特征。选择3D-ResNeXt-101作为3D主干，以充分利用远程运动信息



- 将二维CNN和三维CNN按通道拼接而成作为输入
- 使用2个2d 卷积提取特征，得到初步融合的特征矩阵A
- 将A的后两维向量化，得到F
- 用F乘F的转置计算Gram矩阵G
- 对G进行softmax计算，生成attention map M
- 将得到的attention map对原feature map进行指导，即计算 $F1 = M^T \cdot F$
- 将之reshape回 $C \times H \times W$ 的维度，得到F2
- 将原始的输入特征图B与F2进行相加即attention模块的输出 $C = \alpha \cdot F2 + B$
- 最后再使用2个2d卷积得到最后的输出

# 项目优势

基于深度学习的复杂状态离岗检测

·搜集到的数据集包含众多多视角监控视频，合理训练后应能较为有效解决多视角的问题

·处理速率到达25FPS，员工检测、员工岗位动作识别较为顺畅，到达实时计算

