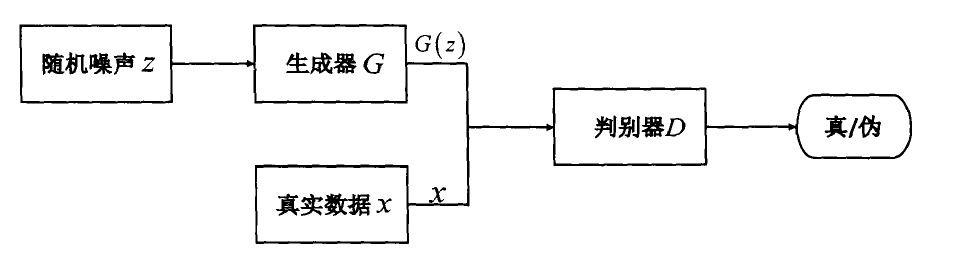
# 智能视频识别技术初步方案

## 视频增强

在视频识别、视频监控等任务中，环境质量对视频识别有着很大的影响。本系统所使用的场景环境复杂，煤矿井下粉尘大、湿气重，矿灯干扰光纤变化等原因，会导致视频对比度降低、细节模糊、质量下降，信息损失严重，这样对视频的利用率大大降低。可以利用GAN强大的生成能力，将复杂环境下的煤矿井下视频映射到虚拟白天正常环境域。考虑增加前景、背景之间的对比度和保持原始视频帧间的稳定性。



## 3D卷积方案

传统卷积神经网络一般是对视频的每一帧图像分别利用2D卷积神经网络来进行识别。这种方式没有考虑时间维度帧间信息。考虑时间维度信息，3D卷积比2D卷积更加有效。用3D卷积核去对输入视频提取特征，如图1所示。把3D卷积应用于广泛使用的Alex、MobileNet、ResNet20等网络。AlexNet模型如图所示。

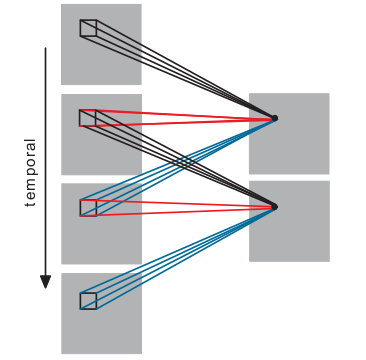


图1 3D卷积

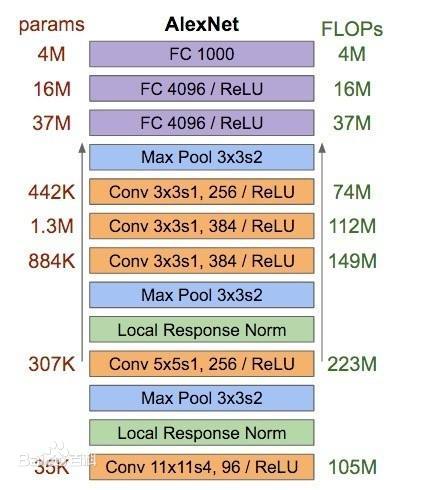


图2 AlexNet卷积

## LSTM方案

LSTM是RNN的一个变种模型，LSTM模型适用于处理时间序列相关的问题。视频帧间有着较强的时间序列信息。这里给出一种基于残差结构、基于注意力的双向LSTM结构网络。图4展示了这种残差注意力机制模块。LSTM优势在于可以全局联系视频的每个时间段，而注意力机制优势在于找自己想关注的部分。

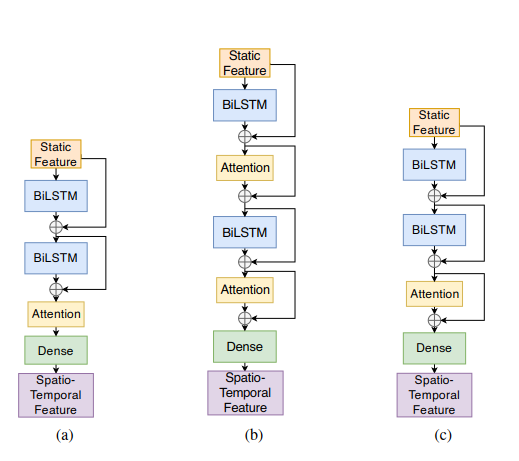


图3 残差注意力机制模块

## 双流方案

考虑视频本身的空间和时间两个维度信息。可以设计两个卷积神经网络，一个卷积网络用语提取视频的空间维度信息，另一个卷积神经网络用于提取视频的时间维度信息。网络有两个输入，一个是用于提取空间信息的单帧图像，另一个是用于提取时间信息的多帧光流，多帧光流即空间运动物体的像素运动的瞬时。这里我们给出一种时空关注的双流协同学习方法TCLSTA，网络模型架构图4如图所示。利用CNN和LSTM对时空注意力进行联合建模，通过视频显著区域定位、以及判别性帧选择。提取学习动静态显著运动特征，对这些特征进行协同优化和自适应学习可以提高视频识别的准确率。

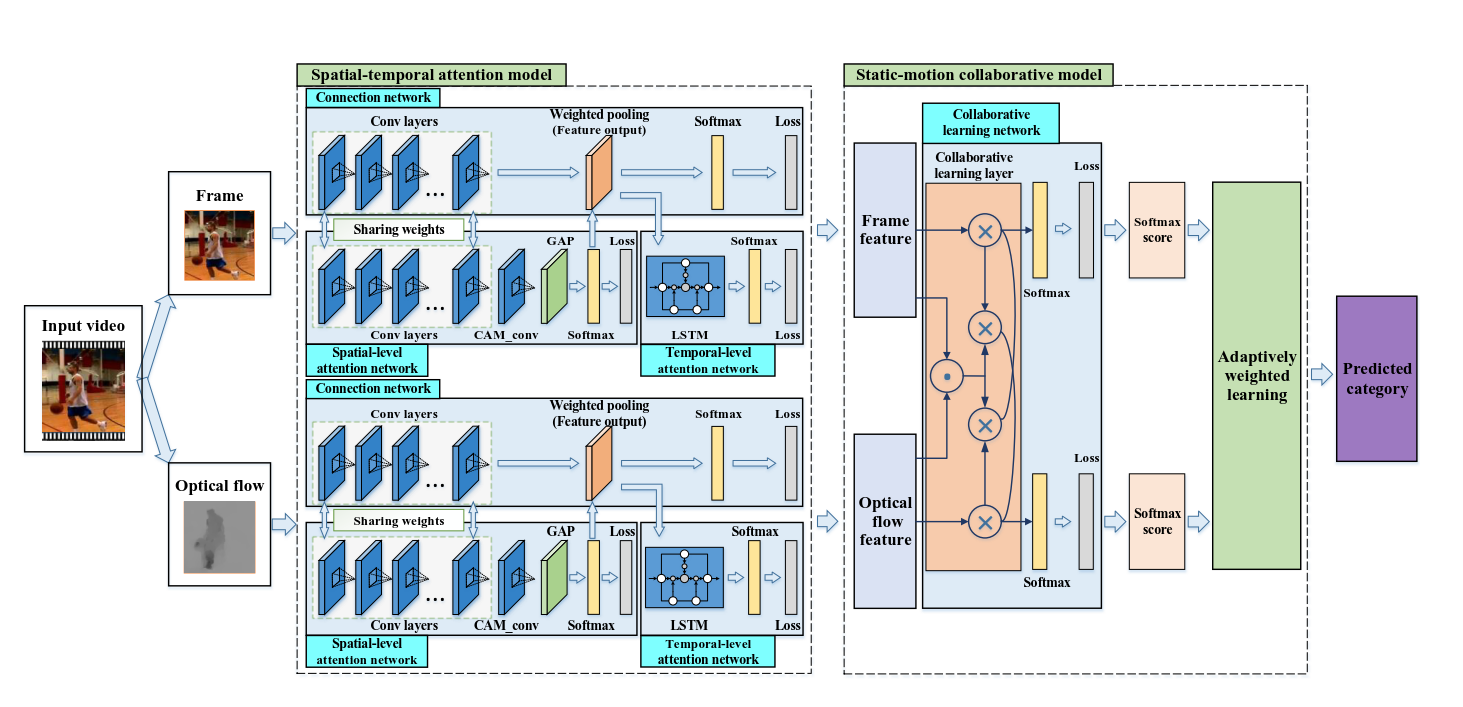


图4 TCLSTA网络结构图