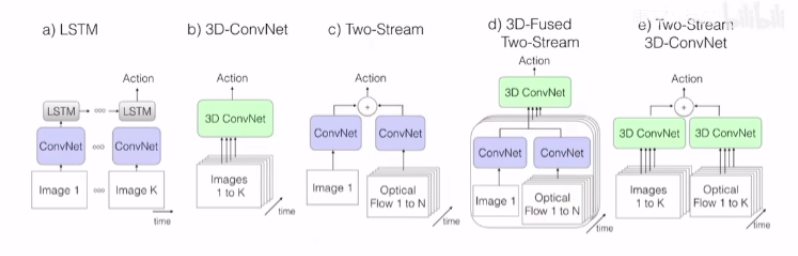
Note: I3D 2018

Sinkoo

April 3, 2022

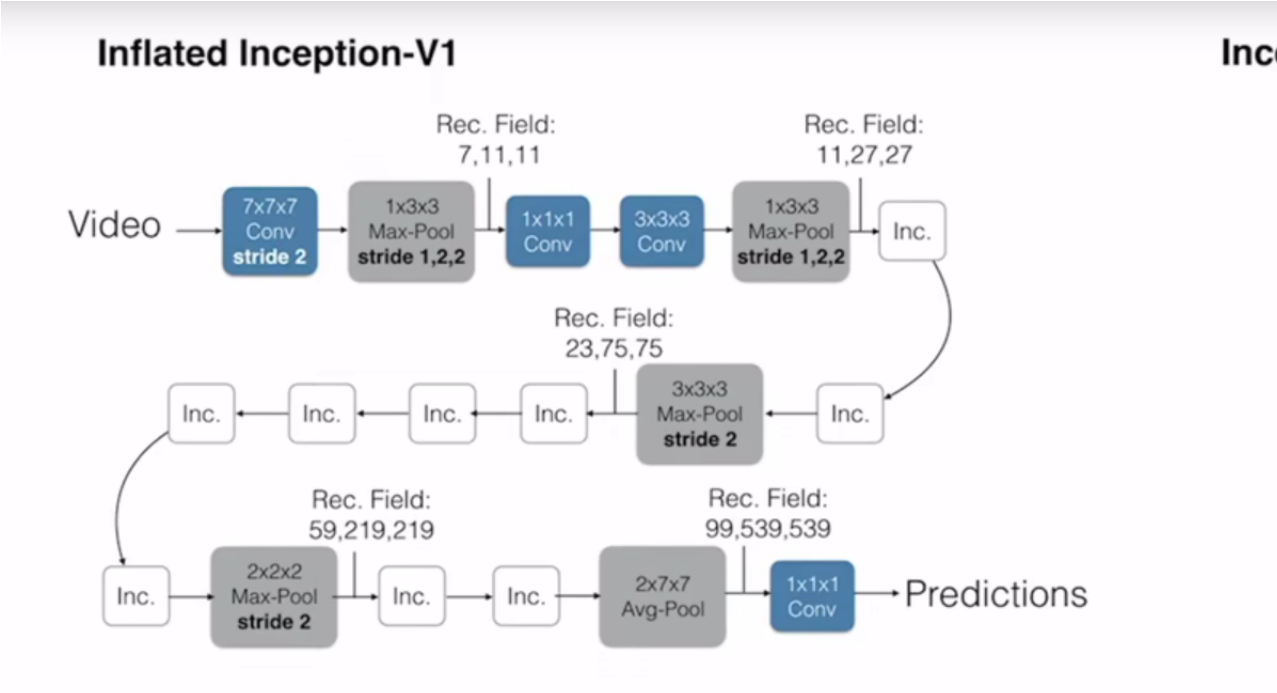
Provided a new model： Inflated 3D convNet（I3D） and a new dataset : Kinetic Dataset (videos, balanced categories, 300000 videos with 400 classes and 400 clips for each, authors use it to pretrain model).
inflation: Example: kernel size: 3*3 $\rightarrow$ 3*3*3.
Video architectures: LSTM(out of date), 3D-ConvNet, Two-Stream, 3D-Fused Two-Stream, Two-Stream 3D-ConvNet(I3D).



1 Method

Inflation (ResNet) Keep the architecture and just change the conv-kernels and pooling layers from 2D to 3D.

Bootstrapping Construct a "Boring" video from a single image by simply repeating it. Use the image to train a 2D CNN and the video to train the 3D CNN, then compare the output. Then we can use ImageNet to pre-train the 3D model.



The Inflated Inception-VI architecture

2 Summary

Transfer Learning can get quite good result when applying to videos(classification). And kinetic dataset helps a lot in video.

3 Other knowledge

[1], optical flow: "https://blog.csdn.net/qq_41368247/article/details/82562165" Optical flow uses the change of pixels through time to get information about previous frames and present frame.
[2], two-stream: "Two-Stream Convolutional Networks for Action Recognition in Videos(NIPS2014)" the model uses 2 Network to process space(image) and time(optical flow) information respectively.