Chapter 1

RNN ConvNet

Local motion

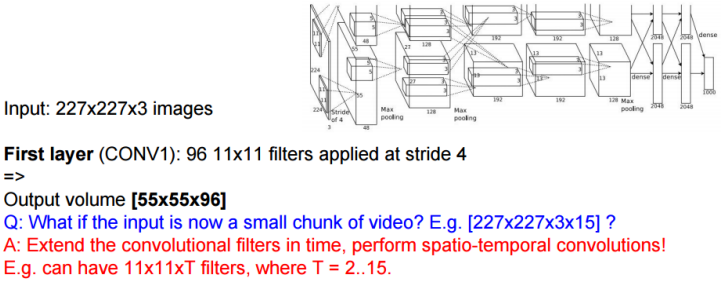

FIGURE 1.1: Getting there

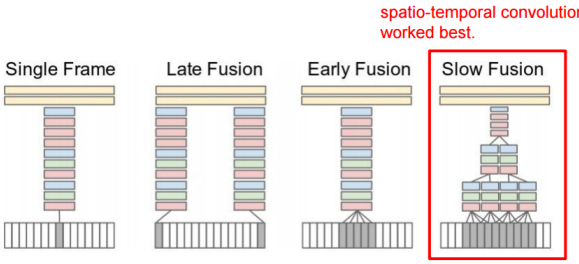

FIGURE 1.2: Large-scale Video Classification with Convolutional Neural Networks, Karpathy et al., 2014]

Global motion

Delving Deeper into Convolutional Networks for Learning Video Representations, Ballas et al., 2016

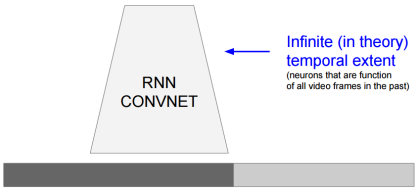

FIGURE 1.3: Large-scale Video Classification with Convolutional Neural Networks, Karpathy et al., 2014]

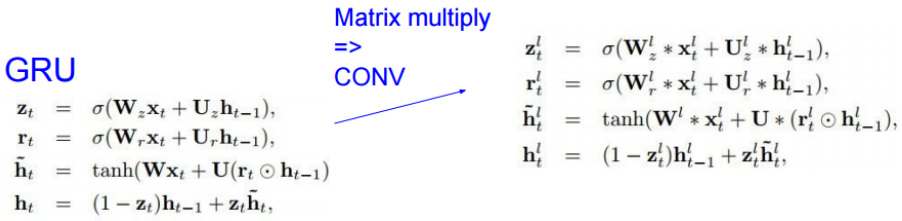

FIGURE 1.4: The idea is to modify GRU adding convolutions. The Gated Recurrent Unit (GRU) is a simplified version of an LSTM unit with fewer parameters. Just like an LSTM cell, it uses a gating mechanism to allow RNNs to efficiently learn long-range dependency by preventing the vanishing gradient problem. The GRU consists of a reset and update gate that determine which part of the old memory to keep vs. update with new values at the current time step.

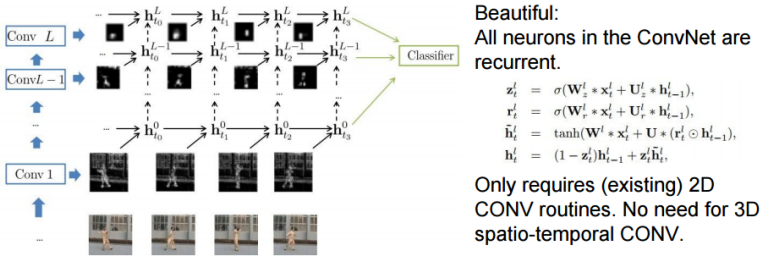

FIGURE 1.5: Original paper diagram

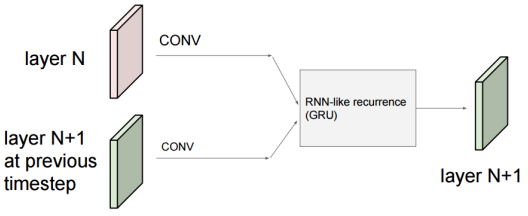

FIGURE 1.6: A more clear diagram

Summary

1. You think you need a Spatio-Temporal Fancy VideoConvNet
2. STOP. Do you really?
3. Okay fine: do you want to model:
   - local motion? (use 3D CONV), or
   - global motion? (use LSTM).

4. Try out using Optical Flow in a second stream (can workbetter sometimes)

5. Try out GRU-RCN! (imo best model)