

Turtles all the way down
with video classification.

Greg Newman

MSc Student Applied Mathematics Stellenbosch

(Ben Herbst & Willie Brink)

HealthQ

- Features are both spatial and temporal.
- Video doesn't always focus on the class.
- Possible to have multiple classes per video.
- Dealing with large datasets is actually really fun.

UCF-101



Playing violin



Handstand pushups



Breaststroke



Brushing teeth

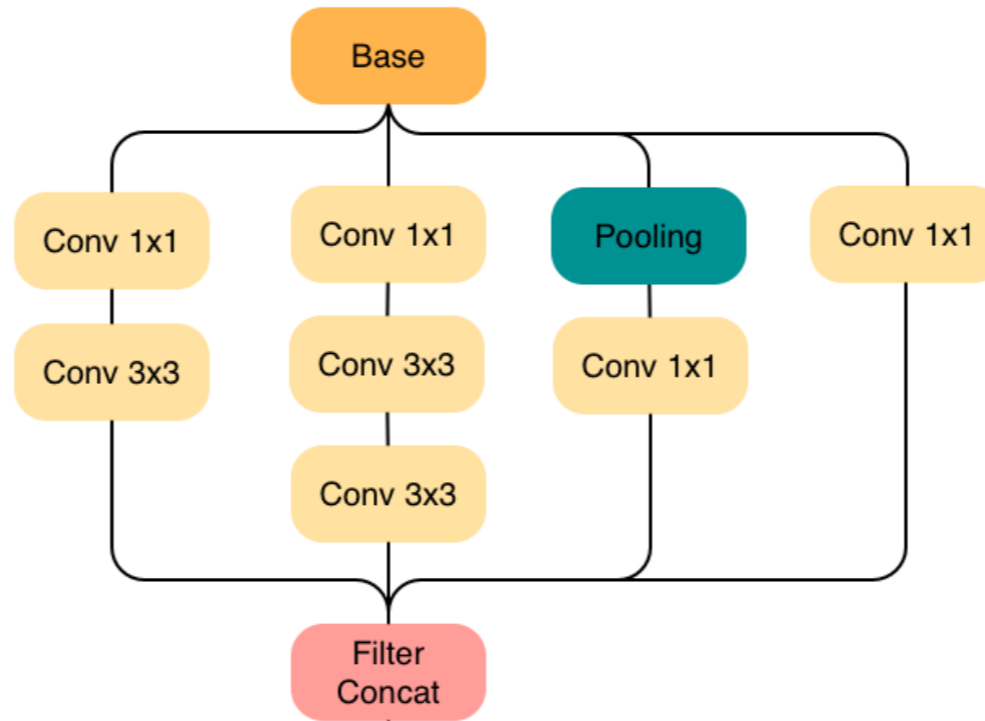


Handstands

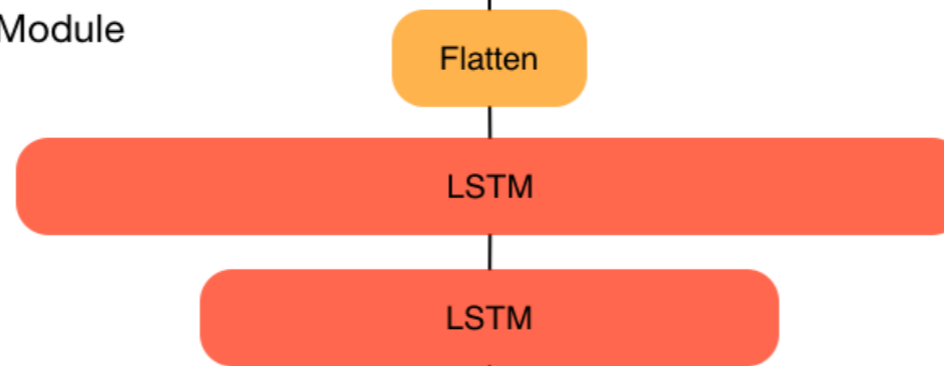


Front Crawl

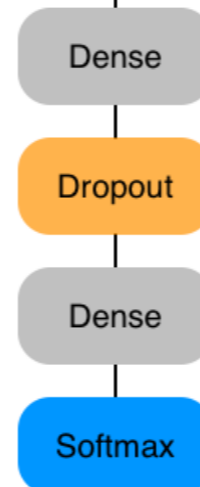
Inception Module



Recurrent Module



Classification Module



Training

- Use pre-trained weights from something like ImageNet.
- Isolate and train the convolutional layers with individual frames.
- Freeze those weights and train the LSTM with features learning from the convolutional layers.
- End-to-end fine tuning.

Why pre-train?

