

Action Recognition

一：相关定义

行为识别（Action Recognition）任务是从视频剪辑（2D帧序列）中识别不同的动作，其中动作可以在视频的整个持续时间内执行或不执行。行为识别是图像分类任务到多个帧的拓展，聚合来自每帧的预测。

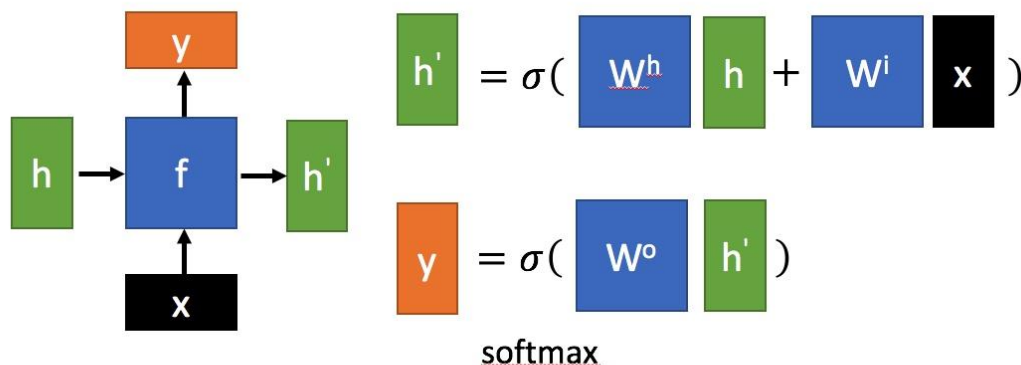
简单的来说就是：对于给定的分割好的视频片段，按照其中的人类行为进行分类。比如女孩化妆、男生打球、跑步等等。

二：相关方法（RNN）

循环神经网络（Recurrent Neural Network, RNN）是一种用于处理序列数据的神经网络。

Naïve RNN

- Given function $f: h', y = f(h, x)$

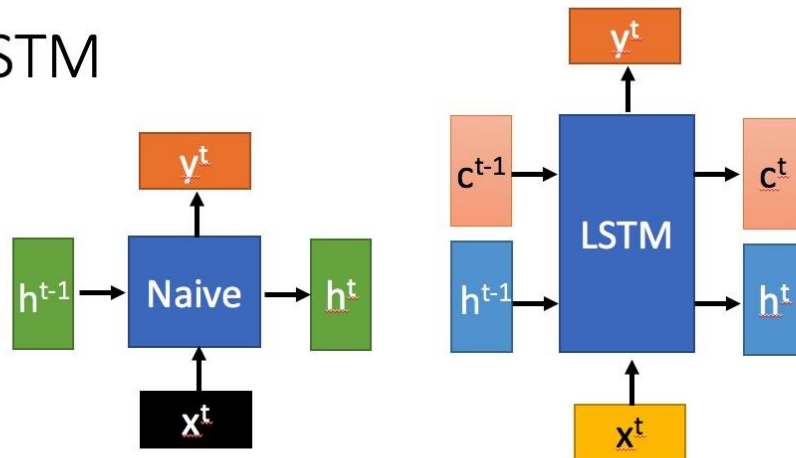


Ignore bias here

二：相关方法（LSTM）

长短期记忆（**LSTM**）是一种特殊的**RNN**，主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题。简单来说，就是相比普通的**RNN**，**LSTM**能够在更长的序列中有更好的表现。

LSTM

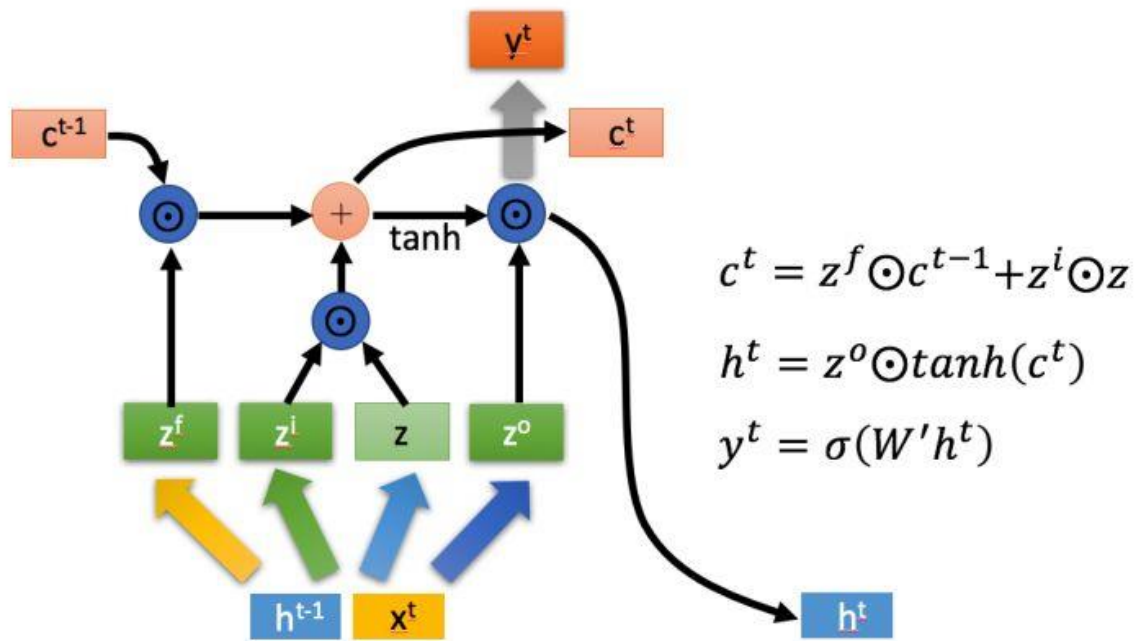


c change slowly $\Rightarrow \underline{c}^t$ is c^{t-1} added by something

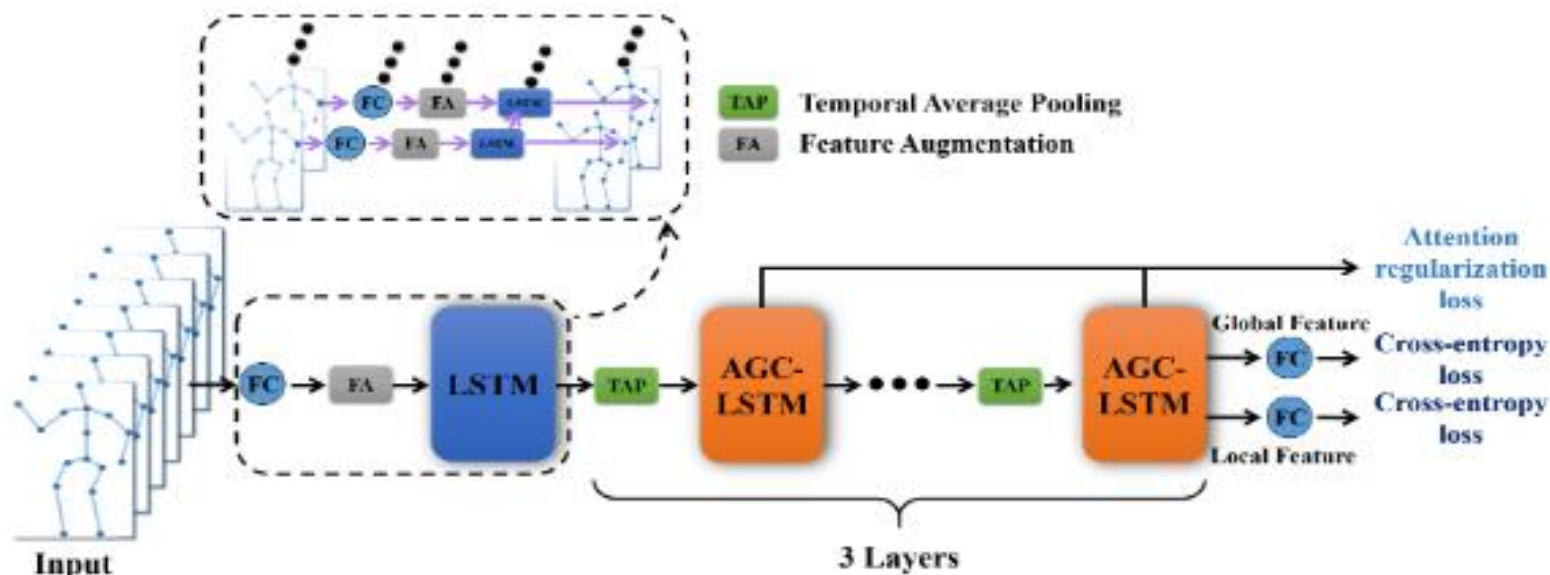
h change faster $\Rightarrow \underline{h}^t$ and h^{t-1} can be very different

二：相关方法（LSTM）

通过门控状态来控制传输状态，记住需要长时间记忆的，忘记不重要的信息



二：相关方法（AGC-LSTM）



- 1, 首先，每个关节的坐标通过线性层转换为空间特征
- 2, 然后将空间特征和两个连续帧之间的特征差连接起来以构成增强特征
- 3, 应用三个**AGC-LSTM**层来建模时空特征。更具体地说，采用注意力机制在每个时间步增强关键节点的特征，这可以促进**AGC-LSTM**学习更多的区分特征。例如，“肘”，“腕”和“手”的特征对于动作“握手”非常重要，应在确定行为的过程中加以增强。
- 4, 最后，我们使用最后一个**AGC-LSTM**层中所有关节的全局特征和聚焦关节的局部特征来预测人类动作的类别。

三：数据集

训练集：“video_data”是一个开放数据集，包含**50**种不同的动作和**3500**个视频剪辑。特别的是，每个文件夹的名称是该文件夹中视频的类标签。

本项目中，需要对视频数据进行预处理。我们提供了一个脚本(video2jpg.py)来实现将视频转换为图片的功能。您可以使用或修改它来满足您的请求

测试集：在“val_data”文件夹中，包含**1500**个不带类标签的视频剪辑。您应该使用您的模型分别按照给定的顺序预测标签，得到一个包含**1500**个元素的列向量，并将预测结果保存为txt格式的标签文件。

四：评分标准

你的成绩= 60% *你的报告得分+ 40% *你的报告准确性得分

最后的成绩不仅仅取决于预测的准确性，也取决于使用方法的创新性

五：相关资源

行为识别相关综述：

- 1, Zhang, H.-B.; Zhang, Y.-X.; Zhong, B.; Lei, Q.; Yang, L.; Du, J.-X.; Chen, D.-S. A Comprehensive Survey of Vision-Based Human Action Recognition Methods. *Sensors* **2019**, 19, 1005.
- 2, Bhoi, A. (2019). Spatio-temporal Action Recognition: A Survey. *ArXiv*, *abs/1901.09403*.
- 3, Junior, V.L., Pedrini, H., & Menotti, D. (2019). Zero-Shot Action Recognition in Videos: A Survey. *ArXiv*, *abs/1909.06423*.

五：相关资源

行为识别相关论文：

- 1, Si, Chenyang and Chen, Wentao and Wang, Wei and Wang, Liang and Tan, Tieniu. An Attention Enhanced Graph Convolutional LSTM Network for Skeleton-Based Action Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2019)
- 2, Majd, Mahshid and Reza Safabakhsh, 'Correlational Convolutional LSTM for Human Action Recognition' (2020) 396 Neurocomputing (Amsterdam) 224
- 3, Donahue, Jeff et al, 'Long-Term Recurrent Convolutional Networks for Visual Recognition and Description' (2017) 39 *IEEE transactions on pattern analysis and machine intelligence* 677
- 4, Du W, Wang Y, Qiao Y. Rpan: An end-to-end recurrent pose-attention network for action recognition in videos[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3725-3734.

五：相关资源

行为相关介绍：

- 1, <https://zhuanlan.zhihu.com/p/100344027>
- 2, 相关论文代码汇总 <https://github.com/jinwchoi/awesome-action-recognition>
- 3, <https://zhuanlan.zhihu.com/p/40964492>
- 4, 骨骼识别和动作生成 <https://github.com/open-mmlab/mmskeleton>
- 5, <https://github.com/HHTseng/video-classification>