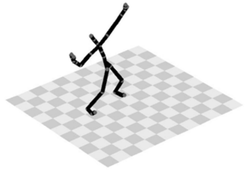
# Regularizing Long Short Term Memory with 3D Human-Skeleton Sequences

Behrooz Mahasseni and Sinisa Todorovic

CVPR 2016

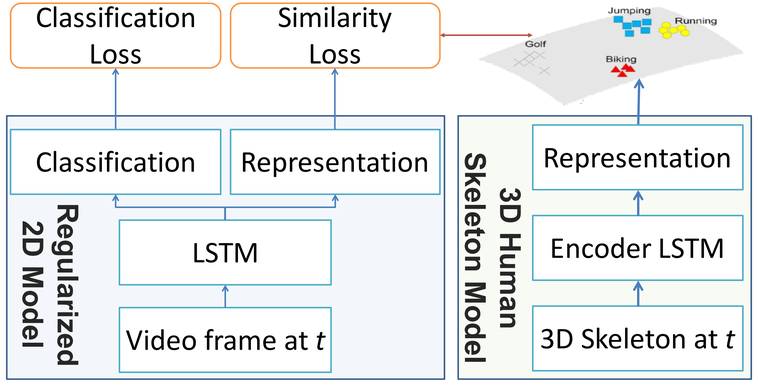
## Problem: Large-scale action recognition

## Key Idea: Additionally train with 3D human-skeleton sequences

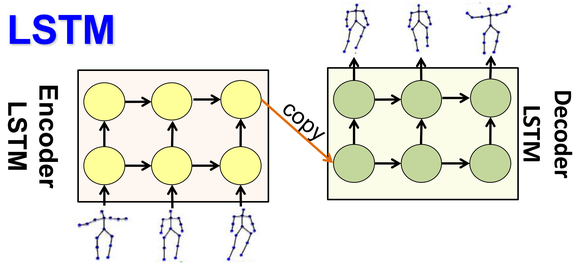


- Instead of typical strategies:
  - Making representations more sophisticated (e.g. deeper)
  - Increasing supervision (e.g. more videos in training)
- 3D data is not available at test time
- Poor coverage of actions in 3D sequences

## Regularized LSTM (R-LSTM)




## Encoder LSTM

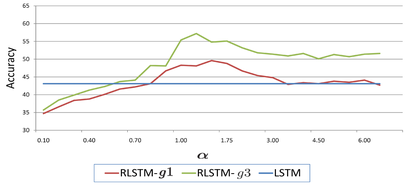



## Training

- Constrains
  - Class aware: $\frac{1}{n_1}\sum_{r_s}|r_s - r_v| - \frac{1}{n_2}\sum_{r_{s'}}|r_{s'} - r_v| \leq 0$
  - Class independent: $\frac{1}{n}\sum_{r_s}|r_s - r_v| \leq \alpha$
- Hybrid backpropagation through time

## Accuracy in [%] on Sports-1M

| Method | Hit@1 | Hit@5 |
|---|---|---|
| CNN | 59.3 | 77.7 |
| LSTM | 71.3 | 89.9 |
| Karpathy et al. '14 | 60.9 | 80.2 |
| Ng et al. '15 | 72.1 | 90.6 |
| Tran et al. '15 | 61.1 | 85.2 |
| R-LSTM | 75.9 | 91.7 |




| Actions | Acc Change |
|---|---|
| Running | 3.7 |
| Badminton | 1.5 |
| Track cycling | 2.2 |
| Road cycling | 1.3 |
| Down hill biking | 0.7 |
| BMX | 0.4 |
| Wind Surfing | -1.2 |
| Fishing | -1.0 |
| Land Surfing | -0.9 |

## Accuracy in [%] on UCF101 & HMDB-51

| Method | UCF-101 | HMDB-51 |
|---|---|---|
| Karpathy et al. '14 | 65.4 | - |
| Srivastava et al. '15 | 75.8 | 44.1 |
| Donahue et al. '14 | 71.12 | - |
| Simonyan et al. '14 | 72.8 | 40.5 |
| Zha et al. '15 | 79.34 | - |
| Tran et al. '15 | 85.2 | - |
| R-LSTM | 86.9 | 55.3 |

**Acknowledgments:** NSF RI 1302700.