

Temporal models

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

Temporal tasks

- Action recognition/
detection
- Action prediction
- Tracking
- Monocular 3D
estimation



Temporal models

- How do we feed videos into CNNs?



Idea 1 – Individual frames

- Treat videos as unordered collection of images
- Ignores time
- Very competitive baseline



Idea 2 – Individual frames + global model

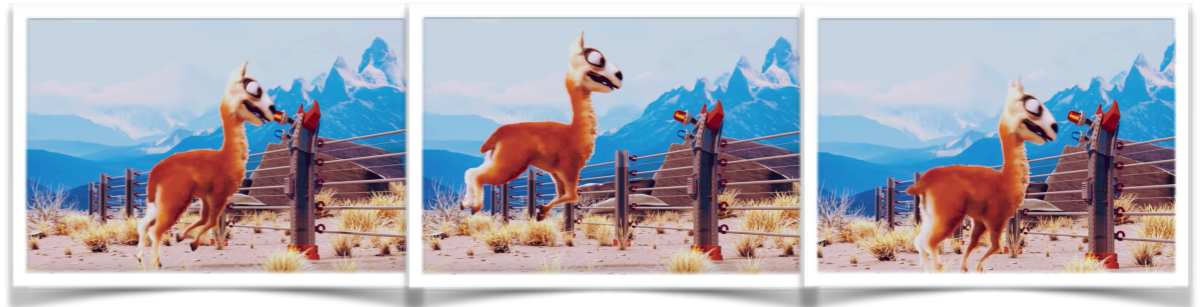
- First parse frames, then model time (e.g. with sequence model)
- Overfits easily
- Hard to train
- Loses motion information



Idea 3 – extend convolutions through time

- Build a 3D convolutional network

...



...

- 2D spatial
- 1D temporal