

Case study: I3D

© 2019 Philipp Krähenbühl and Chao-Yuan Wu

How do we train large 3D CNNs?

- Hard to train from scratch
- Slow to forward and backprop
- No good pre-training task



⋮

Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset, Carreira and Zisserman, CVPR 2017

I3D

- Train image model first
 - on ImageNet
 - as 2D CNN
- then inflate to video model
 - Convert some 2d conv to 3d conv
 - Replicate weights in time



I2+1D

- Train image model first
 - on ImageNet
 - as 2D CNN
- then inflate to video model
 - Add temporal conv
 - Initialize with 1/T



I3D

- Much faster training
 - 2D CNN on ImageNet trains much faster
 - Inflation does not damage network
- Fine-tune on action recognition
 - many days on many GPUs

