Open Problem: Effective temporal operations

Effective temporal operations

- 3D convolutions

  - Inefficient

- 2+1D convolutions

  - Do not care about time much

- Non-local operator

  - Orderless pooling

Non-local Neural Networks, Wang et al., CVPR 2018

Current approaches

- Either, completely ignore time

  - Orderless pooling

- Or apply a brute force hammer

  - 3D convolutions

How should we model time?

- Convolutions

  - Extract local motion

  - No long term tracking

- Motion adaptive convolutions

  - Hard to train or make work

Trajectory Convolution for Action Recognition, Zhao et al., NeurIPS 2018

Should we model time?

- Or is time simply a proxy for space

  - Through ego-motion

- How do we connect time and space?

Image source: EPIC-Kitchens dataset, Damen et al., https://arxiv.org/abs/1804.02748