2+1D convolutions

Factorized 3D convolutions

- Separate spatial and temporal convolutions

  - 2D convolution through space

  - 1D convolution through time
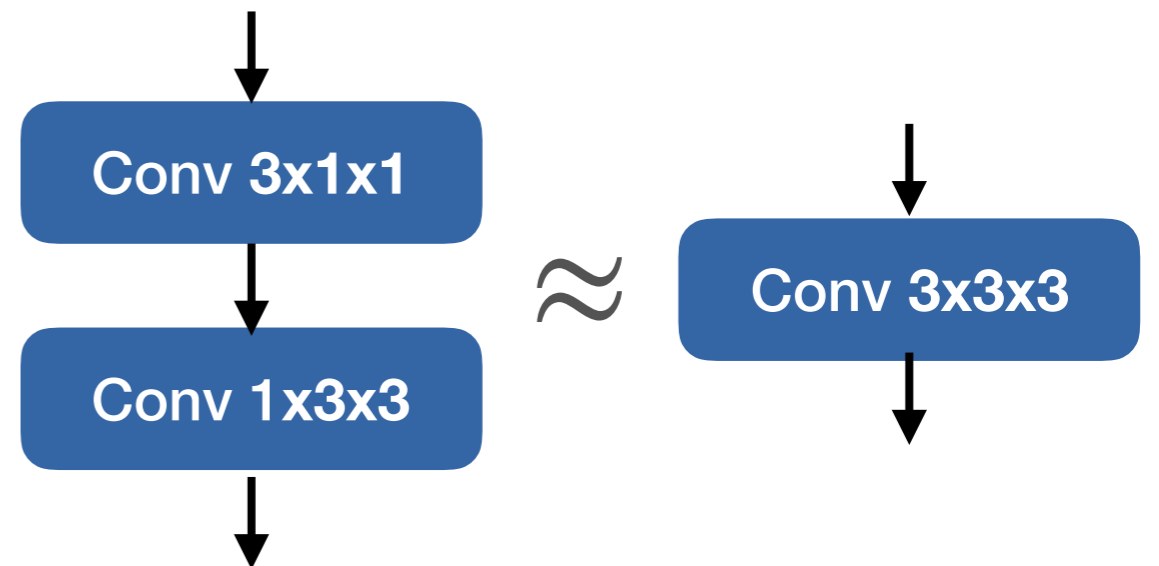
2+1D convolutions

- Fewer parameters

- Less computation

  - $\mathcal{O}((WH + T)C^2)$ vs $\mathcal{O}(WHTC^2)$

- Easier to train