SlowFast Networks for Video Recognition

GxLabs

SlowFast Networks for Video Recognition

Abstract

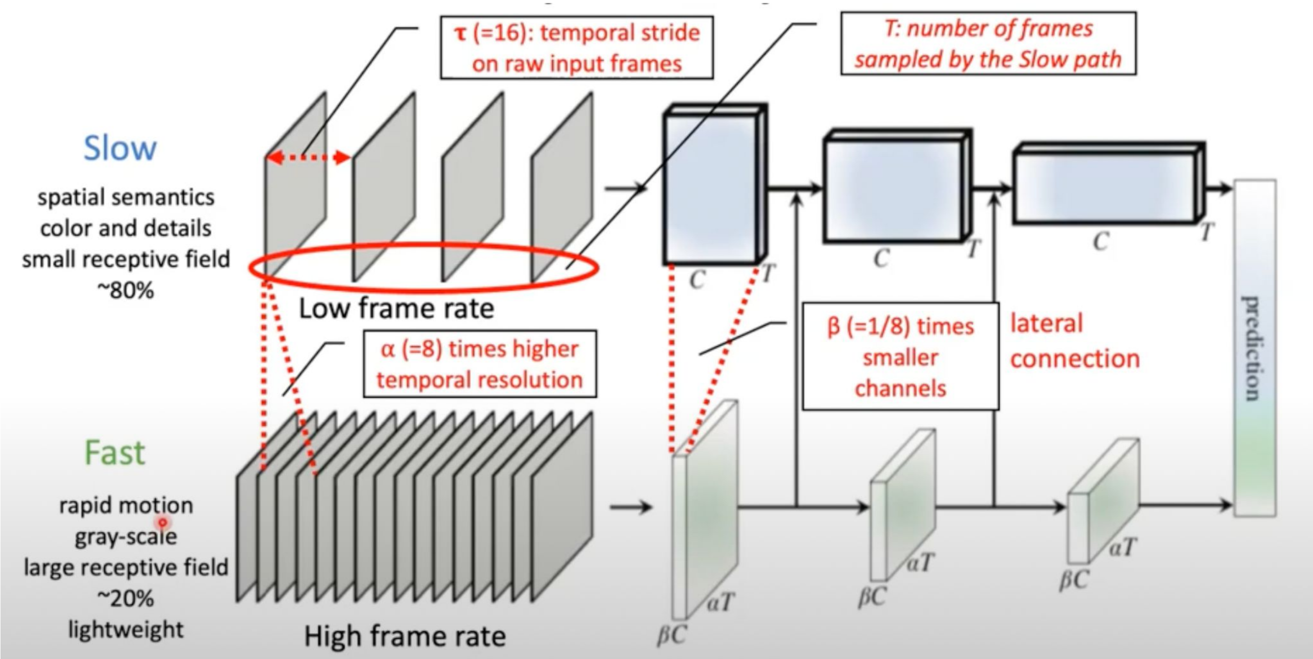
- This method [SlowFast] is partially inspired by biological studies on the retinal ganglion cells in the primate visual system.

- Model involves a Slow pathway and Fast pathway

    - Slow pathway
        - Low frame rate
        - Capturing spatial semantics

    - Fast pathway
        - High frame rate
        - Capturing motion information

SlowFast Networks for Video Recognition



SlowFast Networks for Video Recognition

Lateral Connections

- Attach one lateral connection for every “stage”
  - Right after ResNet $pool_1$, res_2 , res_3 , res_4
  - Unidirectional

- Global average pooling is performed on each pathway’s output
  - Then, Concat -> fully-connected classifier layer

- Feature shape
  - Slow pathway: $\{T, S^2, C\}$
  - Fast pathway: $\{\alpha T, S^2, \beta C\}$

SlowFast Networks for Video Recognition

Lateral Connections

- *Feature shape*
  - *Slow pathway:* $\{T, S^2, C\}$
  - *Fast pathway:* $\{aT, S^2, \beta C\}$

- Time-to-channel: reshape and transpose
  - $\{aT, S^2, \beta C\}$ -> $\{T, S^2, a\beta C\}$
- Time-strided sampling
  - $\{aT, S^2, \beta C\}$ -> $\{T, S^2, \beta C\}$
- Time-strided convolution
  - 3D conv with 5×1^2 kernel
  - $2\beta C$ output channels, stride = a

The output is fused into the Slow pathway by summation or concatenation

SlowFast Networks for Video Recognition

Accuracy/complexity tradeoff

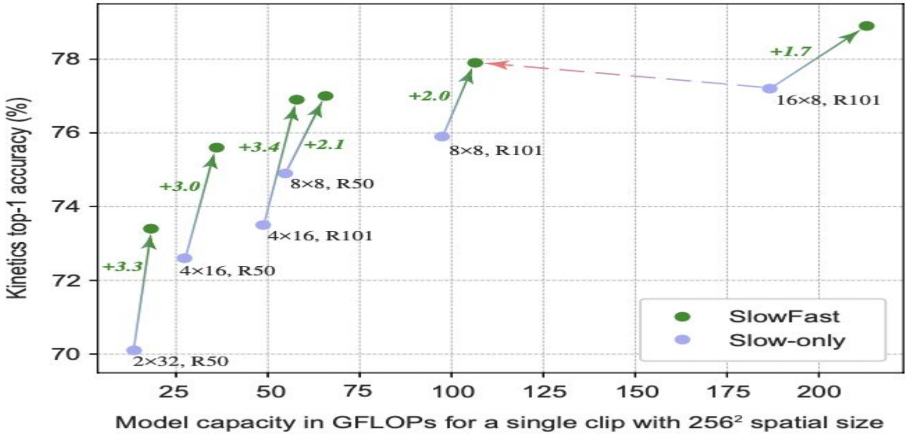


Figure 2. **Accuracy/complexity tradeoff** on Kinetics-400 for the SlowFast (green) *vs*. Slow-only (blue) architectures. SlowFast is consistently better than its Slow-only counterpart in all cases (green arrows). SlowFast provides higher accuracy *and* lower cost than temporally heavy Slow-only (*e.g*. red arrow). The complexity is for a single 256^2 view, and accuracy are obtained by 30-view testing.

SlowFast Networks for Video Recognition

Conclusion

- The time axis is a special dimension
- The SlowFast architecture design focuses on contrasting the speed along the temporal axis
- SlowFast & Two-Stream networks treat space and time differently and share motivation from neuroscience