

① Slowfast video Network.

Slowfast video Network

Slow Path

- ① CNN (3D ResNet)
- ② Large temporal stride. α
(T frames) process
- ③ Process one out of α frames.
- ④ Has higher channel
high $\beta\beta$ computation
- ⑤ (T, s^2, c)

Fast path

- ① Another CNN
- ② Temporal stride \Rightarrow small α/α ; $\alpha > 1$
(αT frames) process
- ③ No temporal downsampling
- ④ Low no of channel
low $\beta\beta$ computation.
- $\beta = 1/\beta$
(weakened spatial ability)
($\alpha T, s^2, \beta c$)

Lateral Connection

- 3 ways
- ① Time to channel $\{ \alpha T, s^2, \beta c \} \rightarrow (T, s^2, \alpha \beta c)$
 \downarrow
 - ② Time Strided sampling $(T, s^2, \beta c)$
 - ③ Time Strided conv. (filter with $(5 \times 1^2, 2 \beta c)$)