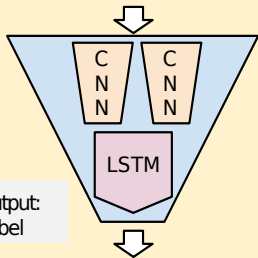
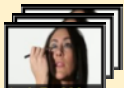


Activity Recognition

Input:
Sequence
of Frames



Output:
Label

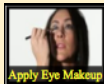
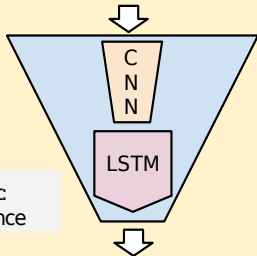


Image Description

Input:
Image

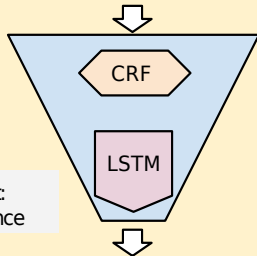


Output:
Sentence

A large building with a
clock on the front of it

Video Description

Input:
Video



Output:
Sentence

A man juiced the orange