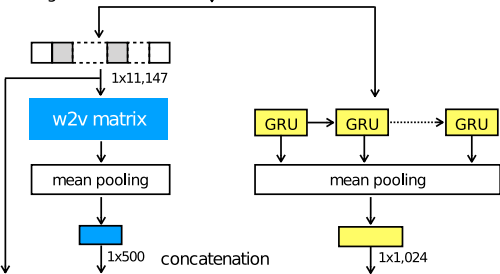


sentence s

a man with beard talking or singing into a microphone

bag of words

sequence of words



Sentence
encoding
network

transformation
network

Video feature space



- v relevant video
- v^- irrelevant video