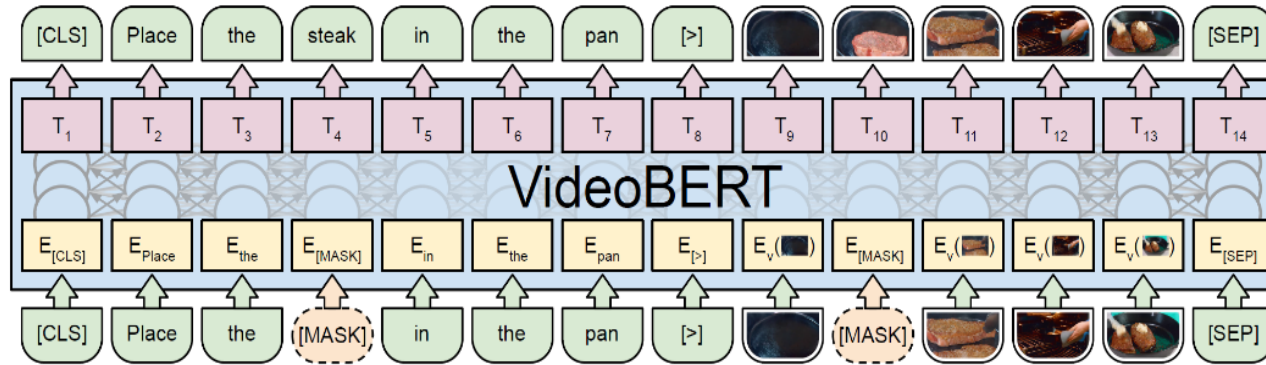


구글 BERT의 정석

VideoBERT, BART

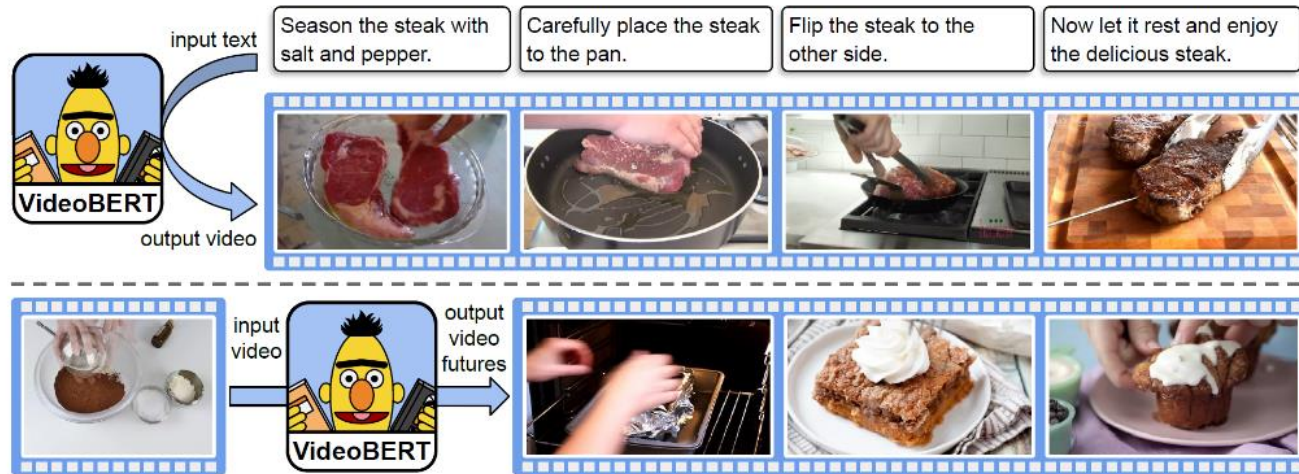
20180376 안제준

VideoBERT 사전학습



- Cloze task와 언어-시각 정렬 task를 이용해 사전 학습 진행
- 사전학습 데이터셋은 요리 비디오와 같은 교육용 비디오 사용
- Cloze task 사전 학습 순서
 - 1. 언어토큰과 시각 토큰(1.5초 구간으로 토큰 생성)을 추출
 - 2. 언어와 시각 토큰을 특수 토큰([>])로 결합
 - 3. 무작위 마스킹
 - 4. VideoBERT에 모든 토큰을 입력
 - 5. VideoBERT에서 반환된 마스크된 토큰의 표현을 분류기(피드포워드 + 소프트맥스)에 입력, 예측
- 언어-시각 정렬 task
 - 언어와 시각 토큰이 시간적으로 서로 정렬되어 있는지를 예측
 - 즉, 텍스트(언어 토큰)가 비디오(시각 토큰)와 일치하는지 여부를 예측
 - BERT의 NSP와 마찬가지로 CLS토큰의 표현을 사용해 언어와 시각 토큰이 서로 정렬되는지 예측

VideoBERT의 응용



- 1. 텍스트-비디오 생성
 - 텍스트가 주어지면 VideoBERT는 해당하는 시각 토큰을 생성 가능
- 2. 다음 시각 토큰 예측
 - VideoBERT에 시각 토큰을 입력해 상위 3개의 다음 시각 토큰을 예측할 수 있음
- 3. 비디오 자막
 - 비디오가 있으면 VideoBERT를 이용해 자막을 지정할 수 있음 (ex. 유튜브 자막(성능은 그닥..))

BERT 라이브러리 탐색

- 이전 장에서 **허깅페이스**의 transformers 라이브러리를 사용하는 법을 배움
- 이번에는 BERT의 라이브러리를 다룸
 - Ktrain 라이브러리
 - Bert-as-service 라이브러리

Ktrain 이해하기

- 케라스 경량 래퍼를 이용해 딥러닝 모델을 더 쉽게 구축, 학습, 배포 가능
- 텍스트 분류, 요약, 질문, 응답, 번역, 회귀 등과 같은 태스크를 더 쉽게 수행할 수 있도록 몇 가지 사전 학습된 모델을 포함
- BERT model을 다양한 down-stream task로 fine-tuning하여 model을 배포하는 형태

Bert-as-service

- 가변 길이 문장을 고정 길이 벡터에 매핑하기 위한 문장 인코딩 서비스

