

A.I. SEE YOU

이미지 기반 시 생성

이미지 기반 시 생성

- **프로젝트 개요:**
 - 연합 학회(투빅스) 9회 컨퍼런스 발표 주제
 - 이미지에 어울리는 시 생성
- **문제**
 - 이미지 – 시의 한국어 캡셔닝 데이터셋 부재
 - 한국어의 특징을 고려한 적절한 모델 구조 설계
 - 시의 특성을 반영한 모델 설계
- **역할**
 - 프로젝트 총괄
 - 이미지 캡셔닝 모델 구현



이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭

- 활용 데이터셋:

한국 관광지 데이터셋

(Korean Tourist Spot; KTS)

- 인스타그램 포스트 수집 데이터
각 샘플은 이미지, 텍스트, 해시태그,
좋아요 수로 구성
- 한국 관광지 10개 레이블에 대하여
각 레이블 당 1,000개 총 10,000개 샘플
- 전수 검사를 통해 개인정보 및 초상권에 대한
비식별화 과정 거침
- <http://ai.dongguk.edu/kts-dataset/>

KTS Images



train/nature-scene/island/images/64.jpg

KTS Text etc.

```
{  
  "img_name": "64",  
  "hashtag": [  
    "#주도",  
    "#드론영상",  
    "#벌설",  
    "#여행"  
  ],  
  "label": "island",  
  "text": "벌설 여행 중 찍은 드론영상 활영~",  
  "likes": "0"  
},  
train/nature-scene/island/island.json
```



train/person-made/palace/images/57.jpg

```
{  
  "img_name": "57",  
  "hashtag": [  
    "#seoul",  
    "#Korea",  
    "#장경궁",  
    "#야간개장"  
  ],  
  "label": "palace",  
  "text": "오늘 사진 느낌 있다!",  
  "likes": "23"  
},  
train/person-made/palace/palace.json
```

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - KTS 데이터셋 전처리
 - KTS 데이터셋의 포스트 내용에 해당하는 'text'의 값을 살펴본 결과 길이가 지나치게 짧거나, 이미지와 관련 없는 내용으로 이루어져 있는 경우가 존재하였음



```
{  
  "label": "restaurant",  
  "image" : "4",  
  "text" : "지루해",  
  "hashtag" : ['#간식', '#바닐라스위트', '#신촌카페']  
}
```

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - KTS 데이터셋 전처리
 - 이미지와 관련된 키워드 정보를 얻기 위하여 clarifai의 Genera Image Recognition Model API를 이용해 이미지 내 키워드 추출하여 ‘keyword’에 추가
 - <https://www.clarifai.com/models/image-recognition-ai>

```
In [2]: from clarifai.rest import ClarifaiApp  
app = ClarifaiApp(api_key=CLARIFAI_API_KEY)
```

```
In [3]: model = app.public_models.general_model  
# response = model.predict_by_url(url='https://samples.clarifai.com/metro-north.jpg')  
  
response = model.predict_by_filename(filename='C:/jupyter_project/tobigs/project/Kore  
an-Tourist-Spot-Dataset-master/kts/total/nature-scene/beach/images/1.jpg')
```

```
In [4]: concepts = response['outputs'][0]['data']['concepts']  
for concept in concepts:  
    print(concept['name'], concept['value'])
```

바닷가 0.9981221556663513

모래 0.9978129863739014

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - KTS 데이터셋 전처리
 - 이미지와 관련된 키워드 정보를 얻기 위하여 clarifai의 Genera Image Recognition Model API를 이용해 이미지 내 키워드 추출하여 ‘keyword’에 추가
 - <https://www.clarifai.com/models/image-recognition-ai>



```
{  
  "label": "restaurant",  
  "image" : "4",  
  "text" : "지루해",  
  "hashtag" : ['#간식', '#바닐라스위트', '#신촌카페'],  
  "keyword" : [  
    '케이크', '크림', '컵(용기)', '초콜릿', '디저트', '커피',  
    '차', '판', '설탕'  
  ]  
}
```

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - KTS 데이터셋 전처리
 - 기존 ‘text’와 ‘hashtag’ 또한 ‘keyword’에 추가한 뒤,
‘keyword’의 단어들을 Khaiii 토크나이저로 분석 후 명사만 추출



```
{  
  "label": "restaurant",  
  "image" : "4",  
  "text" : "지루해",  
  "hashtag" : ['#간식', '#바닐라스위트', '#신촌카페'],  
  "keyword" : [  
    '간식', '카페',  
    '케이크', '크림', '컵', '초콜릿', '디저트', '커피',  
    '차', '판', '설탕'  
  ]  
}
```

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡션ning 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - 시 데이터셋 구축
 - 아마추어 시인 커뮤니티 “시 사랑 시의 백과사전(<http://www.poemlove.co.kr/>)”에서 근현대 및 현대시 14만 건 수집, 전처리 이후 약 7만 4천건 확보
 - BeautifulSoup와 Selenium 이용

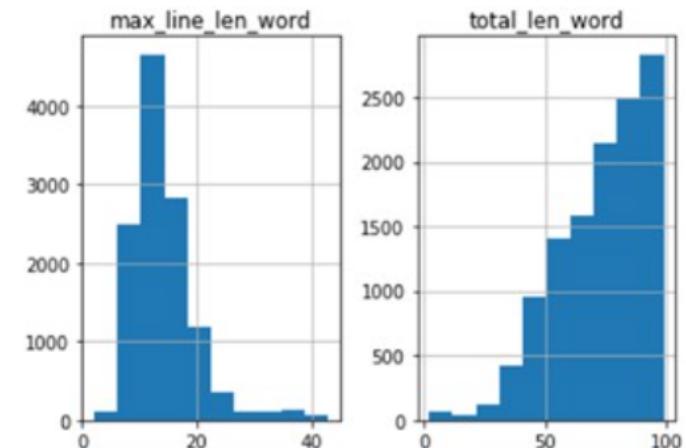
writer	text
0 김소율	★ 봄 가을없이 밤마다 듣는 달도☆ 예전엔 미처 몰랐어요☆ 이렇게 사무치게 그리울 ...
1 김소율	★ 강 위에 다리는 놓였던 것을☆ 건너가지 않고서 바재는 동안☆ 때의 거진 물결은 ...
2 김소율	★ 하다☆ 못해 죽어 달려가 올라좀더 높은 데서나 보았으면한세상 다 살아도살은 뒤 ...
3 김소율	★ 산으로 올라갈까들로 갈까오라는 곳이 없어 나는 못 가오☆ 말마소 내 집도정주작산...
4 김소율	★ 봄 가을 없이 밤마다 듣는 달도☆ 이렇게 사무치게 그리울 줄도☆ 예전엔 미처 몰...
5 김소율	★ 그런대로 한 세상 지내시구로사노라면 잊힐 날 있으리다못 잊어 생각이 나겠지요☆ ...
6 김소율	★ 옛날 우리나라먼 뒤풀의진두강 가람가에 살던 누나는의붓어미 시샘에 죽었습니다☆ 아...
7 김소율	★ 산에는 꽃 피네☆ 꽃이 피네☆ 갈 봄 어를 없이☆ 꽃이 피네☆ 산에☆ 산에☆ 피...
8 김소율	★ 나 보기가 역겨워☆ 가실 때에는☆ 말없이 고이 보내 드리우리다☆ 영변에 약산☆ ...

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - 시 데이터셋 구축 - 전처리
 - 특수문자, 영어, 숫자 삭제
 - 시와 관련 없는 문장 삭제

"['풍경', '불고기 처량하게', '쇠 된 불고기', '하릴없이 허공에다', '자기 몸을 낸다 치네', '저 불고기', '절 집을 흔들며', '맑은 불소리 쏟아내네', '문득 절 집이 불소리에 번지네', '절 집을 불고', '불고기 떠 있네 이 게시물은 님에 의해 시동록없는 시 올리기으로 부터 이동됨']"

- Khaiii 토크나이저를 이용한 토크나이징
- 각 샘플별 문장의 개수와 토큰의 개수에 따른 필터링
 - 짧지도 길지도 않은 보통의 길이의 시 생성을 위해 시행의 개수(문장의 수)와 단어의 개수를 제한함
 - 시의 문장 수가 5 이상, 20미만인 경우만 선택
 - 시의 총 토큰 개수가 20 토큰 이하인 경우 제거



이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - 멀티모달 데이터셋 – 시 매칭
 - 시 데이터셋에서 Khaiii 형태소 분석 결과 명사만을 추출

['당신이 있는 그 곳에도',
'아침이면 햇살이 비치고',
'이곳처럼 새소리가 아름답게 들리나요',
'오늘 같은 아침이면 당신도 한 조각의 빵과',
'한 잔의 커피를 마시며 하루의 시작을',
'배불림으로 행복해 하나요',
'당신이 있는 그곳에도',
'오늘처럼 별이 좋은 날이면',
'아름다운 꽃들이 소망처럼 피어나나요',
...
'적당히 배부를 수 있는 양식과',
'항상 오늘처럼 꽃이 피어 향기 나서']



['곳', '아침', '햇살', '새소리', '오늘',
'아침', '조각', '빵', '잔', '커피',
'하루', '시작', '배불림', '행복', '오늘',
'별', '날', '꽃', '소망', '꽃송이',
'속', '꿀', '송이', '냄새', '치장', '기분',
'나들이', '햇살', '오늘', '요란',
'새소리', '배부', '양식', '오늘', '꽃',
'향기', '행복', '웃음']

이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - 멀티모달 데이터셋 – 시 매칭
 - 멀티모달 데이터셋의 ‘keyword’의 각 명사와 시 데이터셋의 명사에 대해 Pre-train된 FastText의 임베딩 벡터의 총합을 구하여 코사인 유사도가 가장 높은 시를 매칭

```
{  
  "label": "restaurant",  
  "image" : "4",  
  "text" : "지루해",  
  "hashtag" : ['#간식', '#바닐라스위트', '#신촌  
카페'],  
  "keyword" : [  
    '간식', '카페', '케이크', '크림', '컵',  
    '초콜릿', '디저트', '커피', '차', '판',  
    '설탕' ]  
}
```

```
[  
  '곳', '아침', '햇살', '새소리', '오늘',  
  '아침', '조각', '빵', '잔', '커피',  
  '하루', '시작', '배불림', '행복', '오늘',  
  '별', '날', '꽃', '소망', '꽃송이',  
  '속', '꿀', '송이', '냄새', '치장', '기분',  
  '나들이', '햇살', '오늘', '요란',  
  '새소리', '배부', '양식', '오늘', '꽃',  
  '향기', '행복', '웃음' ]
```

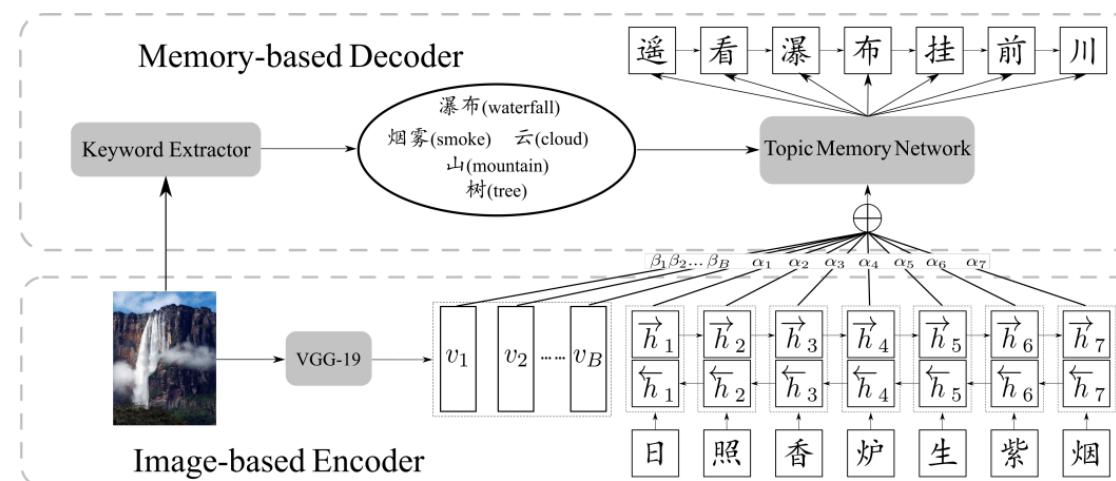
이미지 기반 시 생성

- 문제 1 : 학습 데이터 구축(이미지 - 시 캡셔닝 데이터셋 구축)
- 해법 : 기존 멀티모달 데이터셋의 텍스트 데이터와 시 데이터를 텍스트 유사도를 기반으로 매칭
 - 멀티모달 데이터셋 – 시 매칭
 - 매칭된 시를 ‘text’로 추가하고, ‘keyword’와 ‘text’를 엑소브레인의 Korean_BERT_Morphology 모델 입력에 맞는 형태로 변환

```
{ "label" : "restaurant",
  "image" : "4",
  "keyword" : ['식사/NNG_','생일/NNG_','차/NNG_',
    '우유/NNG_','장/NNG_','이킹/NNG_','커피/NNG_',
    '판/NNG_','아침/NNG_','디저트/NNG_','케이크/NNG_',
    '크래커/NNG_','과자/NNG_','크림색/NNG_','식품/NNG_',
    '컵/NNG_','초콜릿/NNG_'],
  "text" : '[CLS] 밀가루/NNG_ 반죽/NNG_ 을/JKO_ [SEP]
[CLS] 다리미/NNG_ 오븐/NNG_ 에/JKB_ 넣/VV_ 어/EC_ 굽/VV_ 으면/EC_ [SEP]
[CLS] 따뜻/XR_ 하/XSA_ ㄴ/ETM_ 식빵/NNG_ 이/JKS_ 완성/NNG_ 되/XSV_ ㄴ다/EC_ [SEP]
[CLS] 나/NP_ ㄴ든/JX_ 날/NNG_ 마다/JX_ [SEP]
[CLS] 밀가루/NNG_ 으로/JKB_ 반듯/XR_ 하/XSA_ ㄴ/ETM_ 식빵/NNG_ 을/JKO_ 만들/VV_ ㄴ다/EC_ [SEP]'}
```

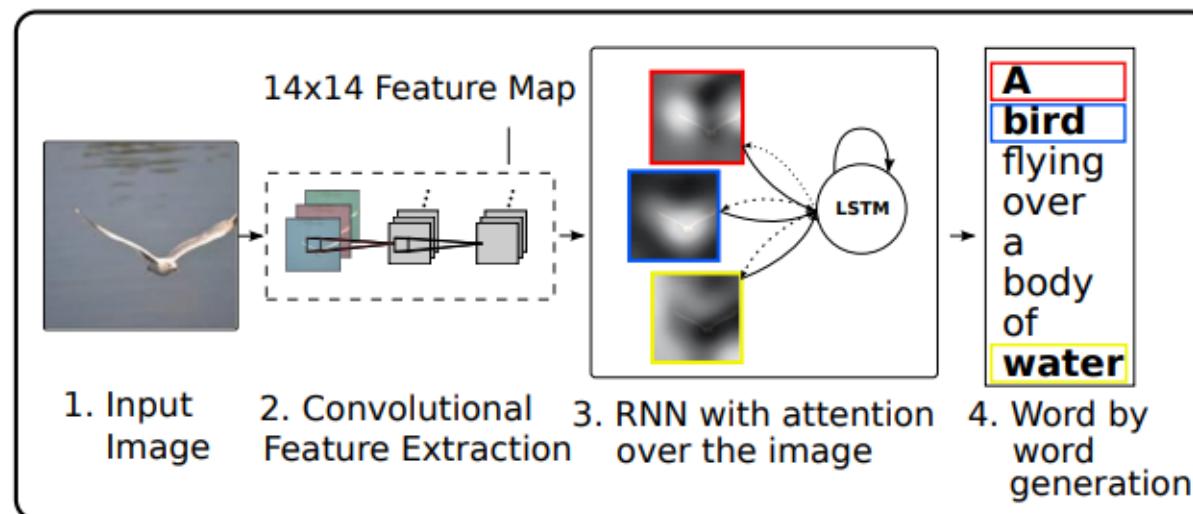
이미지 기반 시 생성

- 문제 2 : 한국어의 특징을 고려한 적절한 모델 구조 설계
- 해법 : 선행 연구로부터 아키텍처 차용
 - 선정 기준 : 프로젝트 기간과 구현 능력을 고려하여 구현이 비교적 단순한 모델 선택
 - Xu, Linli, et al. "How images inspire poems: Generating classical chinese poetry from images with memory networks." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1. 2018.
 - 입력 이미지를 VGG-19로 인코딩한 feature와 Image Recognition API에서 추출한 키워드를 모델의 입력으로 이용 중국어 시 생성



이미지 기반 시 생성

- 문제 2 : 한국어의 특징을 고려한 적절한 모델 구조 설계
- 해법 : 선행 연구로부터 아키텍처 차용
 - Xu, Kelvin, et al. "Show, attend and tell: Neural image caption generation with visual attention." *International conference on machine learning*. PMLR, 2015.
 - 이미지 캡셔닝에 어텐션을 적용한 대표적 연구
 - 벤치마크 모델에서 각 이미지와 토큰 간의 어텐션 스코어 계산의 단순화를 위해 적용

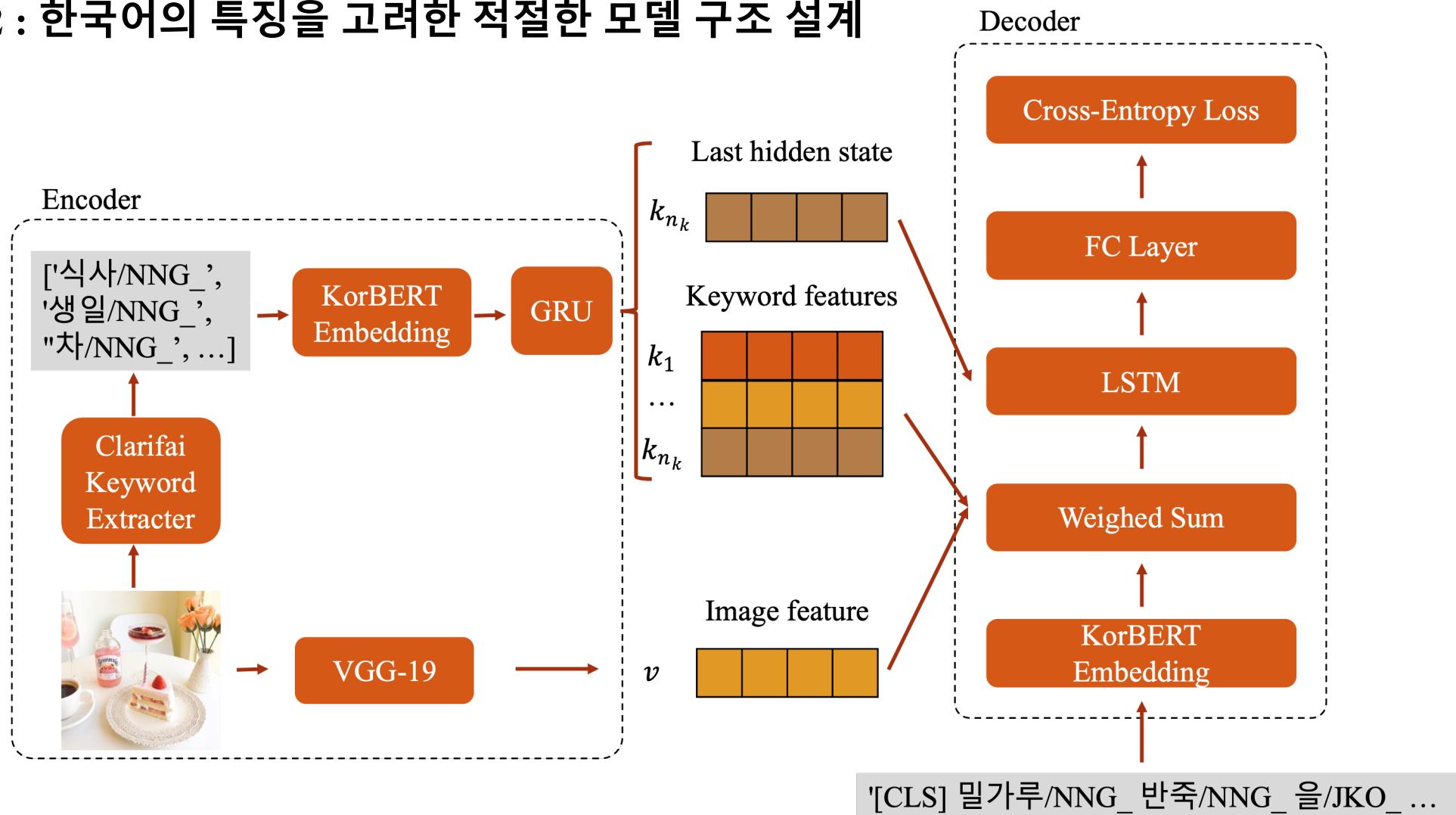


이미지 기반 시 생성

- 문제 2 : 한국어의 특징을 고려한 적절한 모델 구조 설계
- 해법 : 한국어 특징 고려한 변경
 - 중국어는 고립어로, 각 단어의 어순이 중요하며, 매 입력이 하나의 문자이므로 공백 단위로 토크나이징 적용하여도 LSTM을 이용한 문장 생성이 한국어에 비해 어렵지 않음
 - 그러나 한국어의 경우 교착어에 해당하여, 조사, 어미, 접사에 따라 체언과 용언의 문장 성분과 문장 의미가 달라지기 때문에, 단순 공백 단위가 아닌 형태소 단위로 토크나이징하면서, 동일한 문자라 할지라도 문장 성분을 구분해 처리해야 높은 성능을 기대할 수 있다.
 - 이를 위해 ETRI 형태소 토큰을 vocab으로 사용하는 엑소브레인의 KorBERT_Morphology 모델의 임베딩 레이어를 입력으로 활용

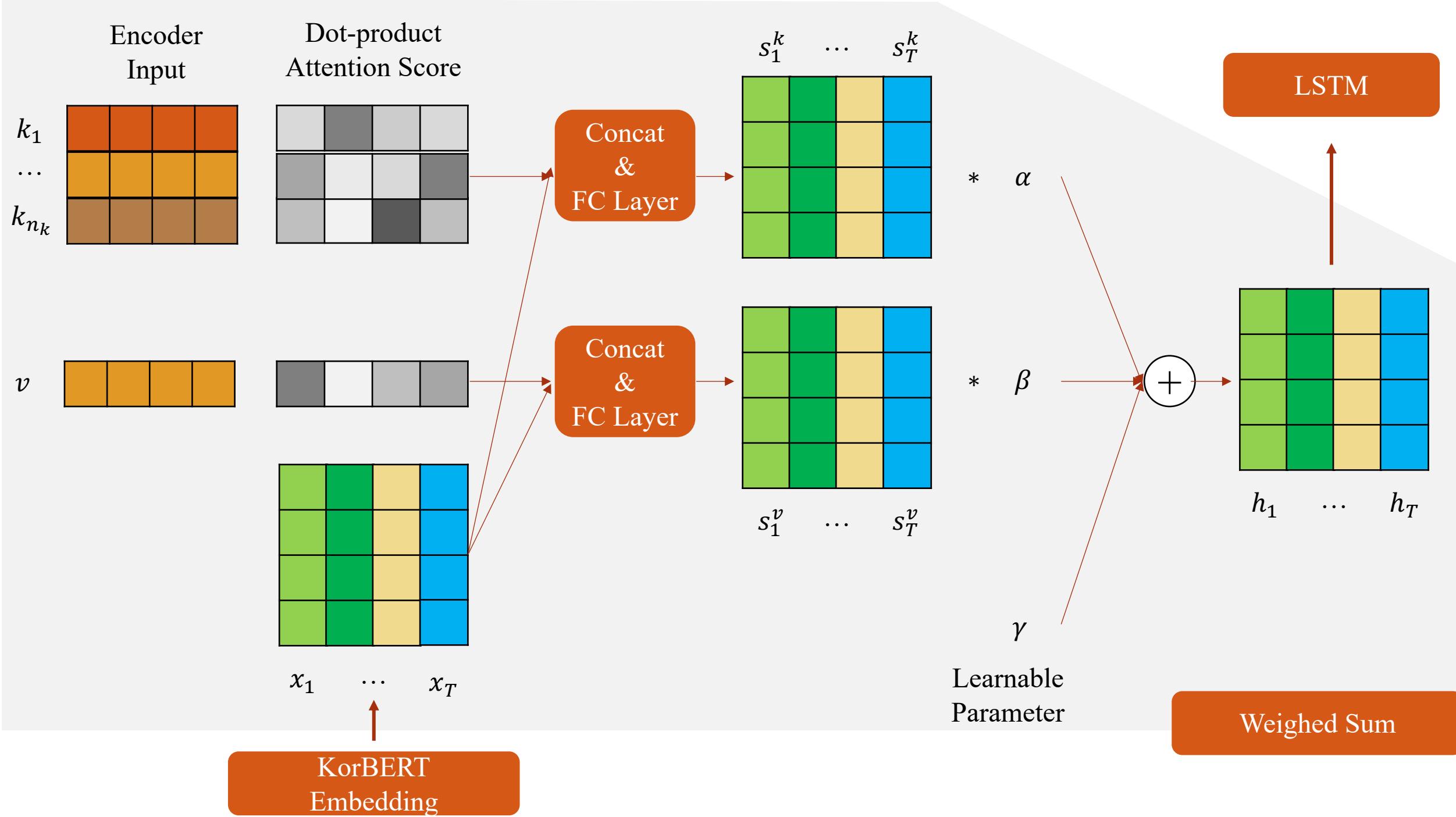
이미지 기반 시 생성

- 문제 2 : 한국어의 특징을 고려한 적절한 모델 구조 설계



이미지 기반 시 생성

- 문제 2 : 한국어의 특징을 고려한 적절한 모델 구조 설계
- 해법 : 시의 특성을 반영한 설계
 - 산문시를 제외한 일반적인 시는 보통의 글과 달리 단락에 의해 개행이 일어나지 않고 시행 별로 개행이 일어나며, 시를 구성하는 시행 및 한 문장을 이루는 단어의 개수가 적은 편
 - 따라서 학습 데이터 전처리 과정에서 이를 고려해 필터링함
- LSTM 모델은 한 번의 truncated BPTT 마다 시 문장 한 줄을 처리하는데 새로운 시행을 처리 할 때마다 키워드 정보를 망각하고 무관한 문장을 내뱉는 경우가 발생
- 따라서 매 Forward 마다 각 토큰에 대하여 맥락 정보를 반영하도록 설계함



이미지 기반 시 생성

- 최종 산출물



'[UNK] 가/VV_ 디가/EC_[UNK][UNK]로/JKB_ 빠지/VV_ 냐다는/ETM_ 말/NNG_ 이/JKS_ ',
'[CLS] [UNK] 이/JKC_ 아니/VCN_ 라는/ETM_ 것/NNB_ 르/JKO_ ',
'[CLS] [UNK] 포/NNG_ 의/JKG_ 매력/NNG_ 에/JKB_ 빠지/VV_ 어/EC_ 보/VX_ 냐/ETM_ 사람/NNG_ 은/JX_ 알/VV_ 냐다/EC_ ',
'[CLS] [UNK] 포/NNG_ 가/JKS_ 얼마나/MAG_ [UNK] 냐/ETM_ 곳/NNG_ 이/VCP_ 냐가/EC_ 를/JKO_ ',
'[CLS] [UNK] [UNK] 고/NNG_ 의/JKG_ [UNK] 냐/ETM_ 풍경/NNG_ 이야/JX_ 말/NNG_ 하/XSV_ 르/ETM_ 것/NNB_ 도/JX_ 없/VA_ 고/EC_ ',
'[CLS] 호수/NNG_ 같/VA_ 은/ETM_ 바다/NNG_ 의/JKG_ 맛/NNG_ 너무/MAG_ 도/JX_ [UNK] 하/XSA_ 고/EC_ ',
'[CLS] [UNK] 에/JKB_ 담/VV_ 은/ETM_ 바다/NNG_ 의/JKG_ [UNK] 맛/NNG_ 또한/MAG_ 천하/NNG_ 의/JKG_ [UNK] 이/VCP_ 지/EC_ ',
'[CLS] [UNK] 섬/NNG_ 의/JKG_ 매력/NNG_ 은/JX_ 더/MAG_ 하/XSV_ 르/ETM_ [UNK] 도/JX_ 없/VA_ 고/EC_ ',
'[CLS] 여러/MM_ 섬/NNG_ 을/JKO_ 오가/VV_ 는/ETM_ 항구/NNG_ 의/JKG_ 배/NNG_ 들/XSN_ 이야/JX_ ',
'[CLS] 꿈/NNG_ 과/JC_ 행복/NNG_ 을/JKO_ 나르/VV_ 느라/EC_ [UNK] 이/JKS_ 없/VA_ 지/EC_ ',
'[CLS] 바다/NNG_ 건너/VV_ 어/EC_ 해/VV_ 지/VX_ 든/ETM_ 풍경/NNG_ 은/JX_ 그야말로/MAG_ [UNK] 다/EC_ ',
'[CLS] [UNK] [UNK] 에/JKB_ [UNK] 이/JKS_ 하나/NR_ 둘/NR_ 켜/VV_ wl/VX_ 르/ETM_ [UNK]',
'[CLS] [UNK] 게/EC_ [UNK] 은/EMT_ [UNK] 같/VA_ 은/ETM_ 태양/NNG_ 이/JKS_ ',
'[CLS] [UNK] 파/JKB_ 함께/MAG_ [UNK] 냐/ETM_ 잔/NNG_ 으로/JKB_ 치/VV_ 는/ETM_ 풍경/NNG_ 은/JX_ ',
'[CLS] 보/VV_ 지/EC_ 않/VX_ 고/EC_ 는/JX_ 그/MM_ [UNK] 모/ETN_ 을/JKO_ 모르/VV_ 냐다/EC_ ',

이미지 기반 시 생성

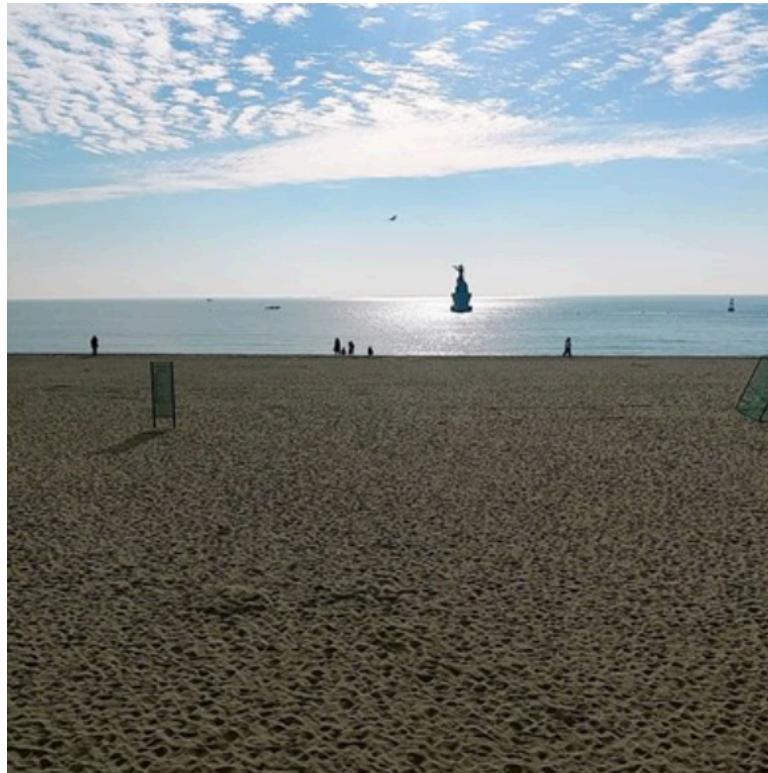
- 최종 산출물



가다가 바다로 빠진다는 말이
[UNK]이 아니라는 것을
[UNK]의 매력에 빠지어 본 사람은 안다
[UNK]가 얼마나 곳인가를
풍경이야 말할 것도 없고
호수 같은 바다의 풍경이야 너무도 [UNK] 하고,
[UNK]에 담은 바다의 맛 또한 천하의 [UNK]이지
[UNK] 섬의 매력은 더할 것도 없고
여러 섬을 오가는 항구의 배들이야
꿈과 행복을 나르느라 [UNK]이 없지
바다에는 [UNK]이 검어지고
[UNK]들도 조용히 집으로 가는 시간
바다 건너 어 해지는 풍경은 그야말로다
[UNK] [UNK]에 [UNK]이 하나 둘 켜질 [UNK]
[UNK]게 [UNK]은 [UNK] 같은 태양이
[UNK]과 함께 [UNK] └ 산으로 지는 풍경은
보지않고는 그 [UNK]ㅁ 을 모른다.

이미지 기반 시 생성

- 최종 산출물



[UNK][UNK] 부터 [UNK]← 마음에
[가슴 설레며 집 [UNK] [UNK] 챙기고서
[UNK] 한 도시와 바쁜 일상을 벗어나아
[UNK][UNK] 유롭게 휴가 떠나는 시기
즐겁ㄴ 휴식 [UNK] 하여보는 여름 [UNK]
맑은 물이 소리내며 [UNK]는 계곡
[UNK]한 바람 불어오는 숲속 [UNK]이든
이른 아침 [UNK]낀 [UNK]의 풍경들이
한폭의 그림처럼 [UNK]는 오토 [UNK]
[UNK] ← 바다 하양ㄴ 펼쳐진 [UNK]
가아볼만한 곳으로 잘 알리어진 [UNK] [UNK] 찾아
[UNK] ← 사람끼리이라면 장소야 어디이든 상관없이
[UNK] 찾아 길떠나는 것 만으로도 너무 행복한
[UNK] 뜻깊은 추억 만들어보는 [UNK] 여름의 절정

이미지 기반 시 생성

- **특징**

- 각 시행의 길이를 평균적인 토큰 개수로 제한하고, CLS 토큰을 이용하여 각 행을 구분하여 일반적인 시의 형태를 갖춘 글을 생성
- 생성된 글에서 ‘바다 건너 어’ ‘벗어나나’와 같이 시에서 자주 볼 수 있는 늘여 쓰는 형태를 관찰할 수 있음
- 생성된 시가 단순 암기의 결과인지 확인하기 위해 수집한 데이터셋에서 포함된 키워드로 검색하여 검증, 2~4 단어 정도 유사한 구는 있었으나 문장이 정확히 일치하지는 않음
- Truncated BPTT 단위인 한 문장을 넘어가는 경우에도 이미지와 키워드, 그리고 이전 시행에 대한 맥락 정보를 잃지 않고 문장을 생성함

이미지 기반 시 생성

- 한계점
 - OOV 문제
 - 형태소분석 기반 언어모델의 OOV 문제
 - Khaiii 대신 해당 모델과 함께 제공하는 형태소 분석 API를 사용하면 개선될 수 있으나 OOV를 근본적으로 해결하진 못함
 - 형태소분석 기반의 언어모델은 교착어인 한국어의 특성을 반영한 모델입니다. 명사/동사에 조사/접미사가 결합된 어절을 의미의 최소단위인 형태소로 구분하여 분석한 언어모델로, 여러 태스크에서 어절 기반 언어모델 보다 우수한 성능을 보입니다. (형태소분석은 본 OpenAPI의 언어분석-형태소분석 API 이용)
 - 한글 자모단위 모델도 시도해보았으나, 형태소 단위의 모델보다 성능이 좋지 못하여 최종 산출물로 선택하지 않음
 - Wordpiece 토크나이저를 활용한 추가 실험 필요
 - 시의 특성을 충분히 반영하지 못함
 - 각운 등 운율 요소를 반영하지 못함
 - 디코딩 방법론
 - 그리디 디코딩 이외에도 빔서치 등 다양한 방법으로 생성 필요

[EOS]
