

MZ 세대를 위한 SNS 사진 문구 & 태그 추천

류혜리, 노가영, 조현지

기획 배경

SNS에 이미지를 업로드할 때, MZ세대 트렌드에 부합하는 문구를 고민하는 어려움을 해소하기 위해서 이미지 캡션 생성 기술을 활용하여 창의적이면서도 MZ세대의 트렌드에 맞는 문구 추천 시스템을 구현하고자 한다

Image Captioning

이미지 캡션은 컴퓨터 비전과 자연어 처리를 결합하여 이미지에 대한 설명을 생성하는 기술이다. 딥러닝 모델을 사용하여 이미지 내의 주요 콘텐츠를 이해하고 이를 자연어로 표현하여 설명을 생성한다.

프로젝트 소개

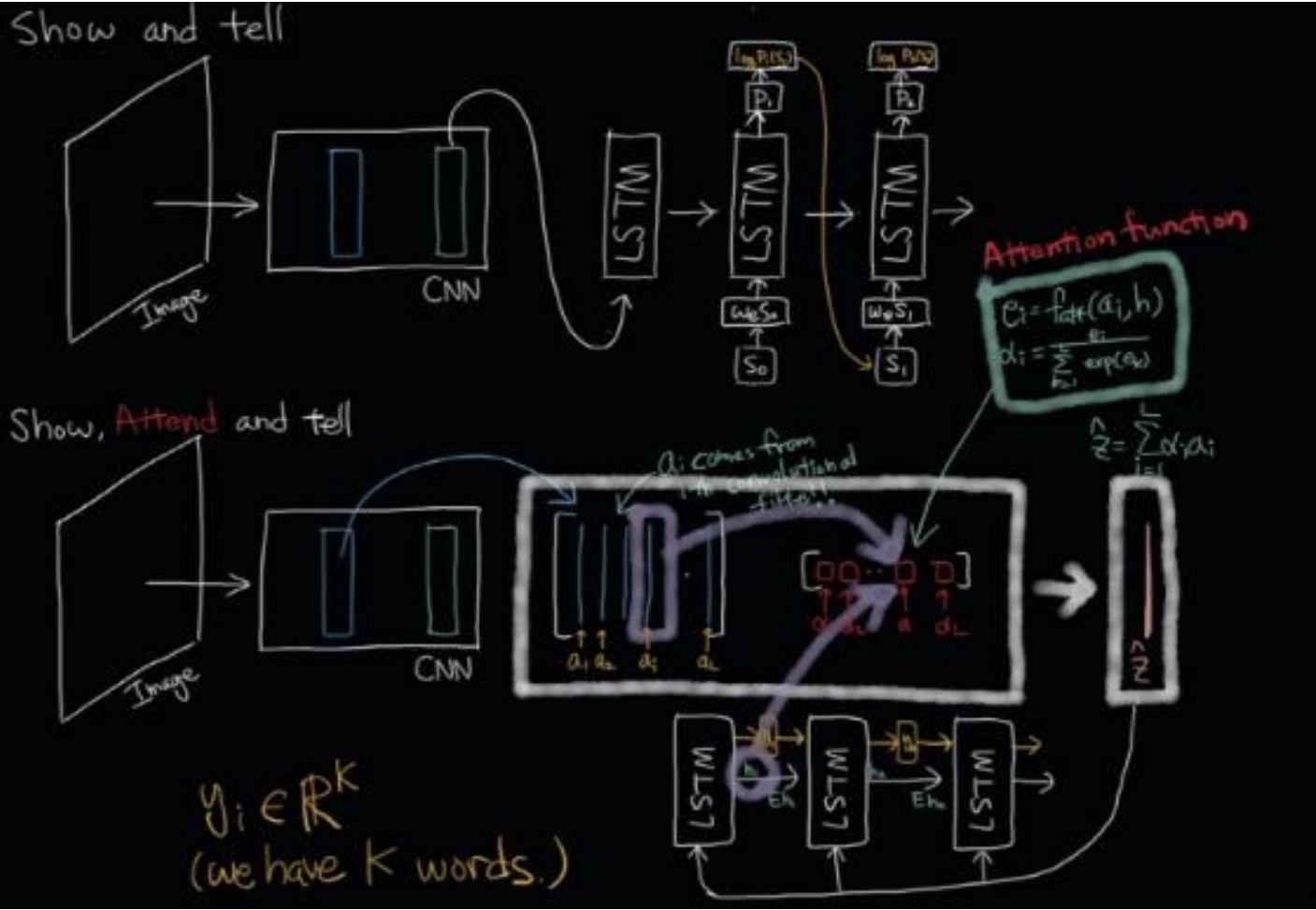
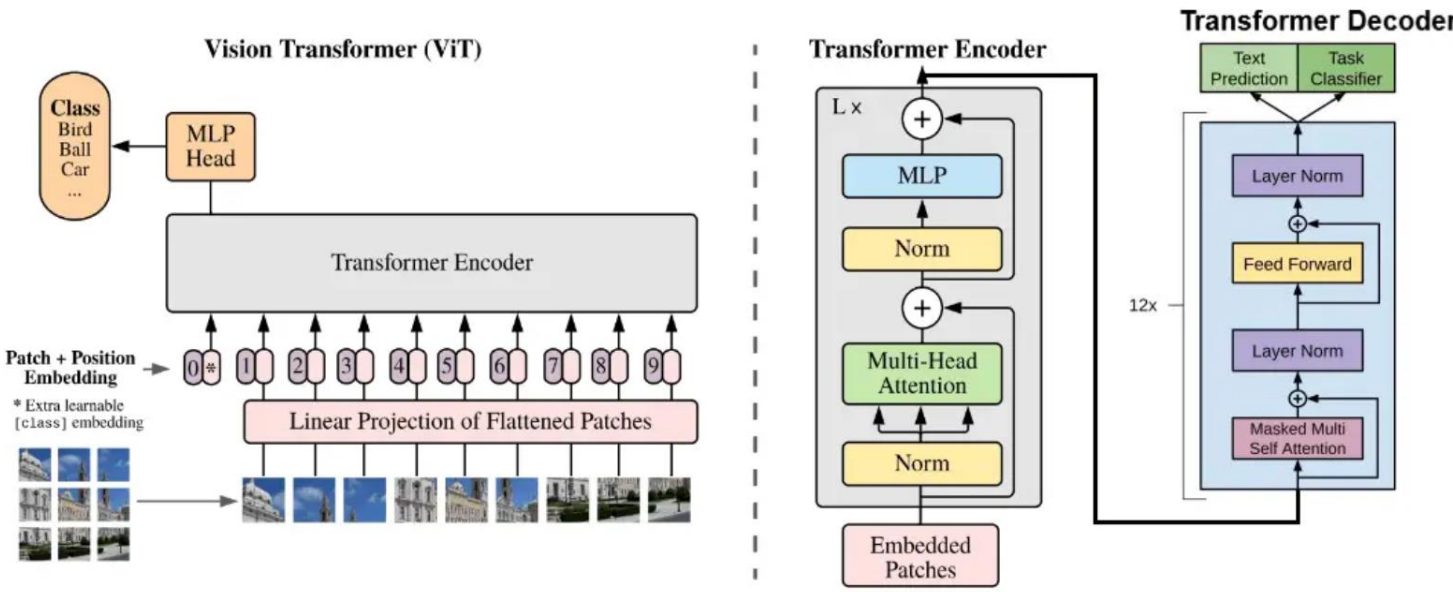
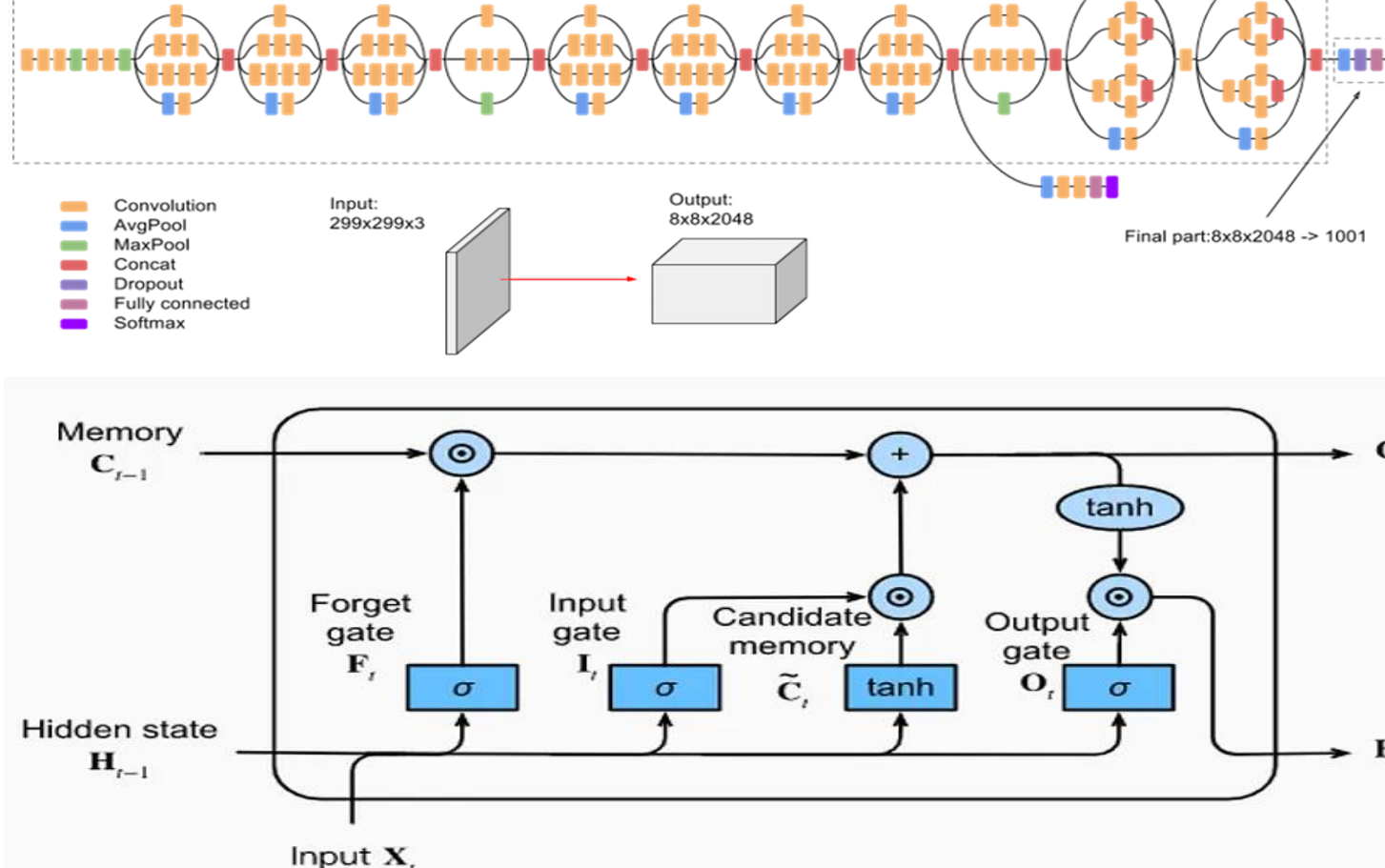
3가지 이미지 캡셔닝 모델을 동일한 데이터셋으로 훈련시켜 성능을 비교분석하고, 각 모델의 출력물인 이미지를 설명하는 영어 문장을 최종적으로 한국어로 번역하기 위해 googletrans API를 활용한다. 최종 출력물을 토대로 투표를 진행하고 최적의 문구 추천 시스템을 선택하고자 한다.

Training Dataset and Test Data(Image)

Flickr8k dataset - 총 8,091개의 이미지를 포함하고 이미지당 5개의 캡션을 가지고 있다.



모델 비교

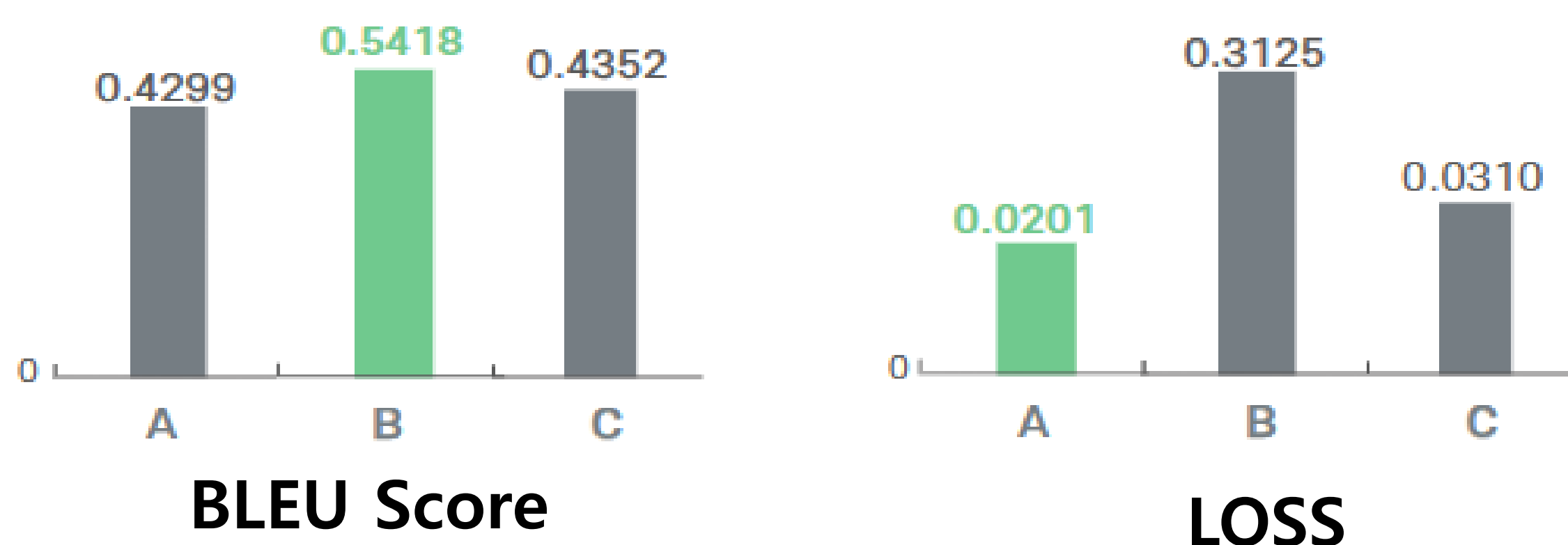
Show and Tell(2015)	Vision Encoder(Swin Transformer) Decoder(GPT2)	InceptionV3 + LSTM
이미지와 문장 간의 시맨틱한 관계를 학습하는 데 초점을 맞추었고, 이미지 특성 추출기는 ResNet-101을 사용하였고, 문장 생성에는 LSTM을 사용	이미지 패치를 나눌 때 self attention computation을 통해 겹치지 않는 여러 조각으로 나누어 정보를 저장하는 Swin 모델과 문장 생성에 GPT2 모델 사용	InceptionV3 아키텍처를 기반으로 한 이미지 특성 추출기와 LSTM 네트워크를 결합한 언어 모델로 구성됨
 <p>논문: https://arxiv.org/abs/1411.4555</p>	 <p>코드 : https://thepythoncode.com/article/image-captioning-with-pytorch-and-transformers-in-python#google_vignette 논문 : https://arxiv.org/abs/2010.11929</p>	 <p>https://www.kaggle.com/code/mrtejas/flickr8k-inception</p>

동일한 Test Data(Image)를 넣었을 때 출력 영어 문장

<start> a white dog is running in the snow . <end>	a dog running through the snow with its tongue out	startseq white dog is running in the snow endseq
----------------------------------------------------	----------------------------------------------------	--------------------------------------------------

성능 비교분석

각 모델을 순서대로 A,B,C라고 지칭하여 그래프에 적용하였다.



BLEU Score는 Vision Encoder 모델이 가장 좋은 성능을 보였고 LOSS는 Show and Tell 모델이 가장 낮은 수치를 보였다.

결과

시간상 제약으로 인해 한국어로 번역하는 부분에 대해 다 담아내지는 못했다. 그러나 test data에 대해 구체적으로, 색깔과 물체의 움직임 등을 잘 표현해낸 것을 알 수 있었다. 다른 여러 지표들이 있었으나, 여기에는 bleu값만 비교하여 나타내 보았다. BLEU는 기계 번역 결과와 사람이 직접 번역한 결과가 얼마나 유사한지 비교하여 번역에 대한 성능을 측정하는 방법이다. 세 모델을 비교하였을 때 B 모델이 가장 성능이 좋았다. 이는 토큰화된 이미지를 자연어로 처리하는 디코더 모델의 변천 (Seq2Seq → transformer → GPT)을 생각할 때, 가장 최신의 모델에서 성능이 가장 좋았음을 알 수 있다.