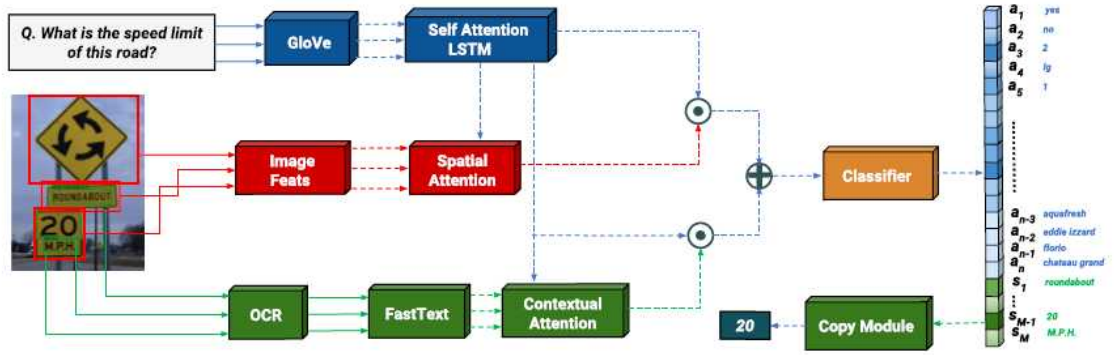
주제: 시각 장애인들을 위한 온라인 쇼핑몰 개발
text & image 에 대한 information extracting model 필요
VQA는 이미지 기반으로 질문에 대답하는 작업

1. LoRRA model



look at the image, reads its text, reasons about the image and text content and then answers. VQA component(이미지와 질문을 이용해 답을 추론), reading component (모델이 이미지에 있는 텍스트를 읽을 수 있게 함), answering component(answer space로부터 예측하거나 reading component를 통해 읽은 부분을 포인트 함)을 포함한다. LoRRA 모델을 위해서는 어떠한 OCR(Optical Character Recognition) model과 attention-based VQA model을 사용할 수 있다.

VQA component
question q의 word를 GloVE에 embed한다. 이를 LSTM을 사용해 encode 한다. fI(v)와 fQ(q)를 통해 attention을 예측하고, weighted average over the spatial feature를 output으로써 제공한다. 이 output과 question embedding을 합쳐서 fVQA를 만든다. fVQA는 answer space에서 각각의 answer가 correct 할 확률을 예측한다.
VQA 2018 challenge winner entry, Pythia v0.3사용

Reading Component
이미지로부터 텍스트를 읽기 위해 OCR model을 사용한다. OCR model이 이미지로부터 word token을 return 한다고 생각한다. VQA component에서의 방식과 비슷한 방식으로 OCR-qustion feature를 만든다.

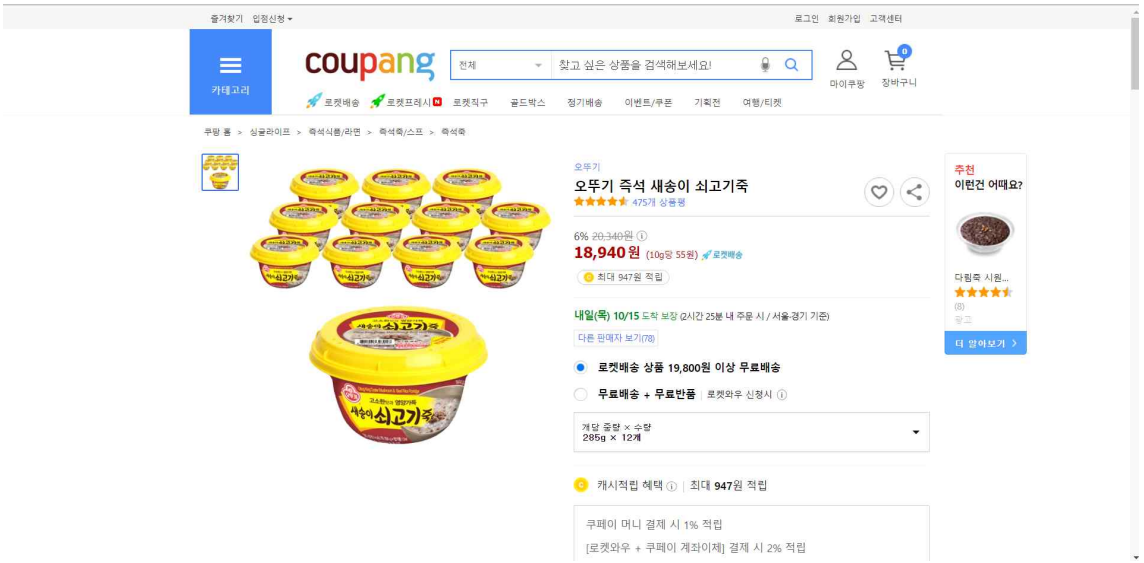
Answer Module
이미지에 있는 텍스트들은 answer space에 미리 정의되어있는 텍스트랑 다른 경우가 많다. 이런 텍스트를 OOV라고 부른다. 이러한 OOV를 reading component를 통해 읽고 answer space에 추가한 후에 추가된 answer에 대한 probability를 예측한다.

VQA 2018 challenge winner entry, Pythia v0.3사용

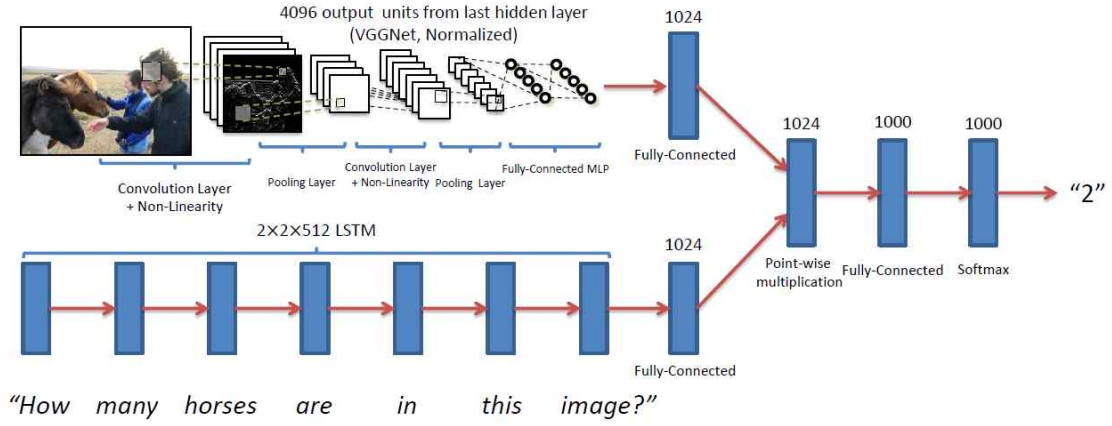
<model>
VQA 2018 challenge winner entry, Pythia v0.3 사용

필요한 데이터셋
text vqa를 위한 쇼핑몰 관련 데이터셋
text based dataset은 많음.
크롤링한 데이터에서 텍스트 뽑아낼 수 있어야함

방법:
text VQA 데이터셋을 통해 VQA model을 training시킨다.
사이트(쿠팡)의 html을 크롤링한 데이터(텍스트)에서 test set (제품의 가격, 수량 등)을 뽑아낸다.
training한 model에 쇼핑몰의 이미지 (쇼핑몰의 이미지는 어떻게 뽑아내는지 확실하게 모르겠음. 실행 시마다 캡처를 할 것인지 혹은 이미지를 얻어오는 것도 크롤링을 통해 할 수 있는 것인지 확실치 않음.)와 질문을 전달한다.
모델이 예측한 값과 크롤링한 텍스트에서 얻은 test set을 비교한다.



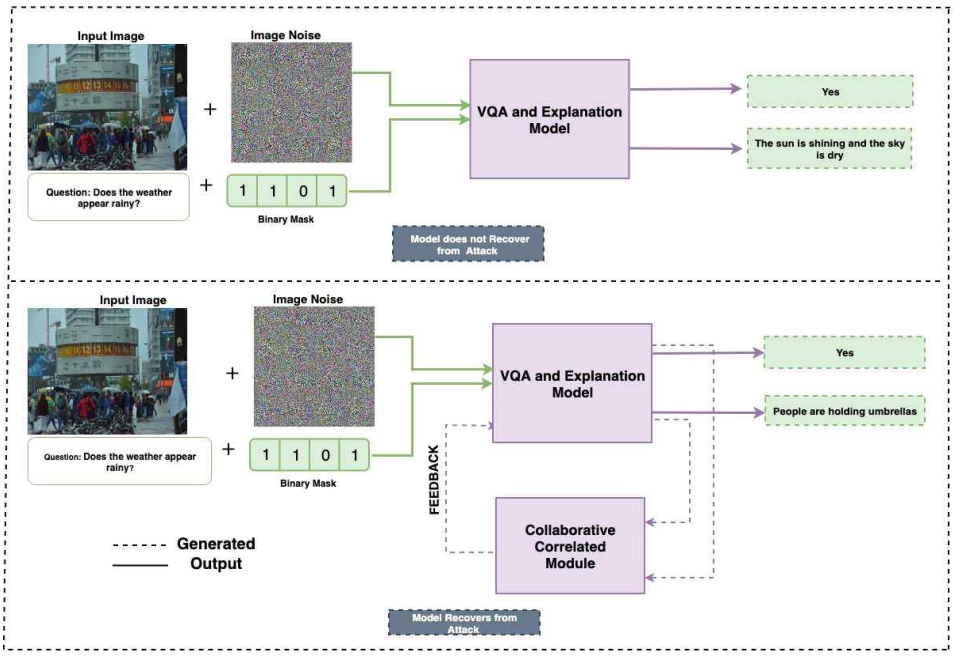
How much is it?
18940 won.

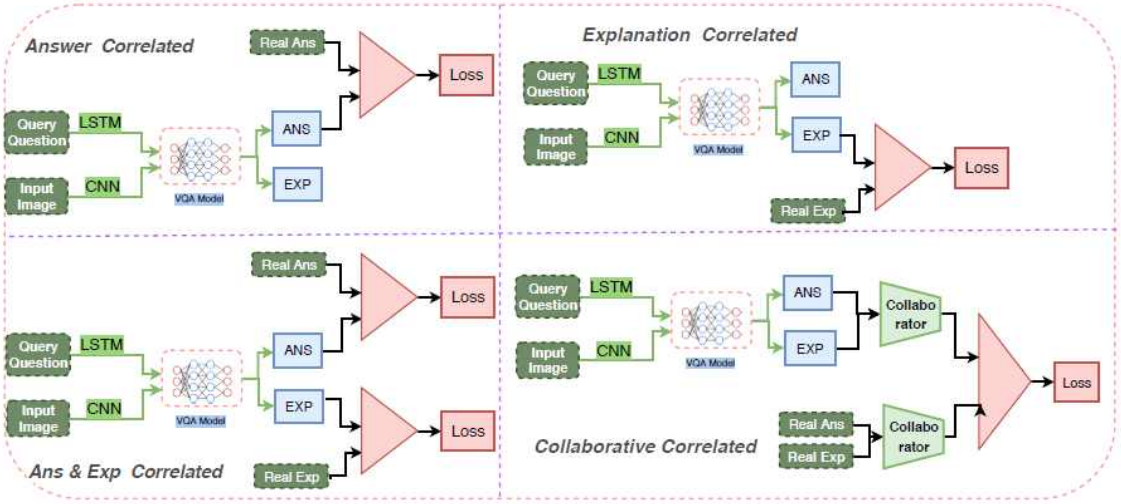


이미지에 대한 VQA를 할 수 있다. 이 모델은 VQA를 위한 기본적인 model로 사용될 수 있을 것이다. two layer LSTM을 사용하여 question과 VGGNet의 마지막 hidden layer를 encode한다. 데이터셋은 https://visualqa.org/ 에서 얻을 수 있다. 이 방법을 통해 가격이나 수량등의 text와 관련된 질문뿐만 아니라 제품의 색상, 형태에 관한 질문에 대한 답도 할 수 있다.

3. 추가적으로 찾아본 noise attack을 막는 model

collaborative correlated network 사용한다. 답과 설명이 정확한지 공동으로 확인한다. 이 방법은 noise attack을 막는데에도 적합하다.




Answer Correlated
Real Ans
Loss
Query Question
LSTM
Input Image
CNN
VQA Model
ANS
EXP
Explanation Correlated
Query Question
LSTM
Input Image
CNN
VQA Model
ANS
EXP
Real Exp
Loss
Real Ans
Loss
Query Question
LSTM
Input Image
CNN
VQA Model
ANS
EXP
Loss
Real Exp
Ans & Exp Correlated
Query Question
LSTM
Input Image
CNN
VQA Model
ANS
EXP
Collabo rator
Loss
Real Ans
Real Exp
Collabo rator
Collaborative Correlated