

# 이미지로 이미지 검색하기

연예인들의 공항 패션은 항상 대중들로부터 많은 관심을 받는다. 편안함을 추구하며 일상적인 모습을 보여주는 사복 패션은 누구나 쉽게 따라 해볼 수 있는 스타일링으로 여겨지기 때문이다. 기사에 실린 사진 속 스타가 입을 옷에 대한 정보를 검색하려고 한다. 어떻게 검색할 수 있을까? '2018년 8월 18일, 설현, 공항 패션', 이런 식으로 연관된 키워드를 입력해서 검색해보니 다행히도 눈썰미 좋은 누군가가 상품의 브랜드를 소개한 글이 블로그에 올라와 있다. 하지만 텍스트로 이미지를 찾는 것은 한계가 있어 보인다. 어떤 경우에는 가능하지만 간혹 원하는 결과가 나오지 않을 때도 많아 이미지를 찾는 데 어려운 경우가 대부분이기 때문이다. 또한 텍스트로 검색해서 얻은 이미지들 중에서 원하는 결과를 찾는 데도 적지 않은 노력이 필요하다.

## 이미지 검색

이미지로 이미지를 검색하는 서비스는 쿼리(query)를 이미지로 받는다. 즉, 검색어를 입력하는 텍스트 기반의 단어 검색이 아닌 이미지를 이용하여 검색하는 방식이다. [그림 1]은 마이크로소프트(Microsoft)에서 제공하는 COCO(Common Objects in Context) 데이터 세트에 포함되어 있는 사진 중 하나이다. COCO는 디텍션(detection), 세그멘테이션(segmentation), 캡셔닝(captioning)을 위한 학습에 활용되는 데이터로, 아래 이미지는 다양한 동물들이 줄지어 있는 박물관 내부의 한 장면을 담고 있다. 사람마다 주목하는 대상이 달라 같은 그림이어도 이를 묘사하는 방식들이 다양하다. 그렇다면 과연 우리는 어떻게 원하는 이미지를 찾을 수 있을까?

[그림 1] 마이크로소프트의 COCO 데이터 세트(data set)\*1

id=55575



a long train of model animals on display to show how noah's arc was filled.  
a few people are standing near a large exhibit of animal statues.  
a large group of wild animals standing next to each other.  
a display of wild animals inside a building.  
a museum display of assorted animals like zebras and giraffes.

\*1 참고 | <http://cocodataset.org/#overview>

컨퍼런스 발표 | 이주영 michael.lee@kakaocorp.com

글 | 이가람 garam24@snu.ac.kr 서울대학교에서 수학과 통계학을 전공하고 최근 머신러닝, 딥러닝, 블록체인에 관심을 갖고 공부하며 그 접점을 찾으려 하고 있습니다. 궁금한 게 많은 문과적인 이과생이고 세상을 조금이라도 더 나은 곳으로 만드는 데 기여하길 바라며 노력하고 있습니다.

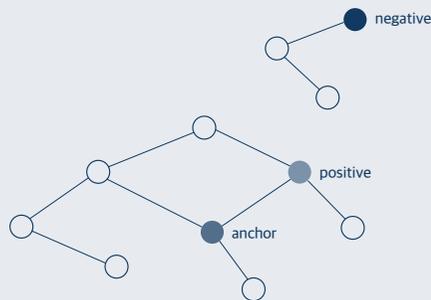
**방법 1) 텍스트로 검색한다**

문장 내에서 주어진 이미지를 가장 잘 표현하는 단어들을 추출한다. 문서를 검색하는 방식과 동일하게 추출된 단어를 사용해서 검색하면 해당 단어와 매핑(mapping)된 이미지를 결과로 얻을 수 있을 것이다. 하지만 이 경우 주어진 사진을 정확하게 묘사하는 문장이 필요하다는 문제점이 가지고 있다. SNS상에 별도의 내용 없이 사진만 올린 경우 검색이 불가능하다. 이를 위해서는 세상의 모든 대상을 인식하여 이미지를 분류하는 분류기가 필요하다. 또한 정제된 표현과 문장의 경우 사진과 직접 연관된 단어를 포함하고 있지만 단어와 사진이 매칭되지 않을 때는 학습에 어려움이 발생한다. 자신의 글을 더 많은 사람들에게 노출시키기 위한 목적으로 노이즈(noise)를 발생하는 단어를 사용했다면 이를 필터링(filtering)하는 과정이 추가적으로 필요하다.

**방법 2) 이미지로 검색하여 일정 거리 내에 있는 유사 이미지를 검색 결과로 제공한다**

각 이미지를 단어와 매핑하여 해당 단어와 일정한 거리에 있는 단어들에 대한 이미지를 검색 결과로 한 번에 노출시킨다. 이를 [그림 2]와 같이 방향성이 없는(undirected) 그래프로 표현할 수 있다. 각 노드(node)들은 이미지를 나타내며 에지(edge)는 각 노드가 연결된 관계를 표현한다. 사용자가 쿼리에 이미지를 입력하면 기준이 되는 노드인 앵커(anchor)로부터 특정 거리 내에 있는 노드들과 같은 그룹 안의 유사한 이미지가 검색 결과로 제공된다. 그래프 내의 노드 중에서 앵커로부터 연결되어 있는 노드는 포지티브(positive)로, 연결되어 있지 않은 노드는 네거티브(negative)로 라벨(label)을 붙인 후 샘플링(sampling)하여 이 세 노드의 관계를 통해 모델을 학습시킨다.

[ 그림 2 ] 이미지를 노드로 매핑한 네트워크 구조



모델은 앵커와 포지티브 노드 사이의 거리가 앵커와 네거티브 노드 사이의 거리보다 더 가까워지도록 학습된다. 이미지 데이터 자체가 매우 큰 경우 함수 F를 사용하여 정보를 압축한다. 또한 세상의 모든 이미지를 학습 데이터로 사용할 수 없기 때문에 새로운 데이터가 들어올 것을 감안하여 일종의 버퍼(buffer) 역할을 할 수 있게 마진(margin) m을 추가한다. 손실함수(loss function)로 트리플렛 함수(triplet function)를 사용하고 합성곱 신경망(convolutional neural network, CNN)을 통해 앵커와 연결된 포지티브 노드는 유사한 이미지로, 네거티브 노드는 다른 클래스의 이미지로 학습한다.

[ 그림 3 ] 학습을 위한 노드 간 거리 비교

$$\begin{aligned}
 & \text{Distance}(\text{Image}_{\text{Anchor}}, \text{Image}_{\text{Positive}}) < \text{Distance}(\text{Image}_{\text{Anchor}}, \text{Image}_{\text{Negative}}) \\
 & \quad \downarrow \\
 & \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Positive}})) < \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Negative}})) \\
 & \quad \downarrow \\
 & \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Positive}})) + m < \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Negative}})).
 \end{aligned}$$

**성능 비교**

다음 표는 스탠퍼드 온라인 프로덕트 데이터 세트(Stanford Online Product data set) 성능을 비교한 결과이다. 학습 데이터는 1만1300여 개의 클래스에 대해 6만 장의 이미지로 이루어져 있다. 성능 측정에 사용한 지표는 Recall@k로 ‘앵커로부터 k번째 가까운 노드까지의 거리 내에 포함되는 앵커의 이미지 개수’와 같은 클래스의 이미지를 갖는 노드 수의 비율을 나타낸다. k가 1인 경우 카카오(kakao)에서 개발한 모델의 성능이 논문에 발표된 가장 좋은 모델의 성능과 비슷하거나 약간 낮은 정도이고, k가 10 혹은 100일 때는 카카오 모델의 성능이 더 좋은 것을 확인하였다.

[ 표 1 ] 스탠퍼드 온라인 프로덕트 데이터 세트를 활용한 성능 비교<sup>\*2</sup>

Stanford Online Product data set [1]  
 120,053 images(train: valid=59,551:60,502)  
 22,634 classes(train: valid=11,318:11316)

	Recall@1	Recall@10	Recall@100
Triplet, Semi-Hard [1]	66.7	82.4	91.9
Lifted Struct [2]	62.5	80.8	91.9
Proxy NCA Loss [3]	73.7	-	-
Sampling Matters [4]	72.7	86.2	93.8
DAML [5]	68.4	83.5	92.3
kakao	73.3	86.2	93.9

\*2 참고 | [http://cvgl.stanford.edu/projects/lifted\\_struct](http://cvgl.stanford.edu/projects/lifted_struct)

### 서비스 응용

‘이미지로 이미지 검색하기’ 서비스는 카카오톡 대화방에서의 이미지 검색 기능(데모버전)에 활용되거나, 이미지를 통해 상품을 검색하는 쇼핑 서비스 그리고 유사 상품을 추천해주는 서비스 등에 응용되고 있다.

[ 그림 4 ] 유사 상품 추천 서비스

서비스 응용(1): 우린 비주얼이 되니까! 이미지 검색!



서비스 응용(2): 다음 쇼핑



 <이미지로 이미지 검색하기>  
브런치로 연결되는 QR 코드입니다.