

# 딥러닝을 이용한 물체인식 기반 상세 영상정보 검색

김선우, 호요성  
 광주과학기술원 전기전자컴퓨터공학부  
 e-mail : {seonwookim, hoyo}@gist.ac.kr

## Detailed Image Information Retrieval based on Object Recognition using Deep Learning

Seonwoo Kim, Yo-Sung Ho  
 Gwangju Institute of Science and Technology (GIST)

### 요 약

기존의 영상검색 알고리즘은 해당 물체가 가지고 있는 일반적인 속성(색상, 무늬, 기하학적 구조정보)을 조합하여 특징벡터를 구성하고, 이 특징벡터와 입력영상과의 매칭을 통해 영상속의 물체의 위치와 그 물체의 일반 분류정보만을 제공했다. 그러나 검색자는 개괄적인 분류정보보다는 해당 물체의 세부적인 정보를 필요로 하는 경우가 많다. 본 논문에서는 먼저 딥러닝 기법에 의해 물체를 인식한 다음, 그 결과를 인터넷 검색 색인으로 활용하여 인터넷으로부터 해당 물체에 대한 상세 정보를 제공하는 알고리즘을 제시한다. 첫 단계인 물체검출 및 인식을 위해 최근 널리 사용되고 있는 YOLOv2 알고리즘을 적용했고, 세부정보는 인터넷 검색결과 얻어진 해당 물체의 설명에 사용된 텍스트를 분석하여 얻었다. 실험 결과 특징이 없는 일반 물체에 대해서는 기존의 물체 검출 결과 이상의 정확도를 보여준 못했지만, 특이한 모양이나 특성이 포함된 영상에서는 기존 방식에 비해 보다 상세한 정보를 얻을 수 있었다.

### 1. 서 론

영상 속에 들어 있는 물체에 대한 상세한 정보를 알아내는 것은 오랫동안 컴퓨터 비전 연구의 중요한 테마였다. 최근에 들어 영상에 포함된 물체를 인식하기 위해 CNN(Convolutional Neural Network) 아키텍처를 활용하는 다양한 알고리즘이 제안되면서 물체인식에 있어서 획기적인 전환점을 맞게 되었다. 그러나 이 경우에도 학습된 물체의 일반적인 속성에 기반을 둔 개략적인 물체정보만을 제공하기 때문에 상세정보의 제공이라는 면에서 개선할 여지가 많이 남아 있다. 예를 들어, 기아자동차의 K5 영상이 입력되었을 때 그냥 “자동차”라고 분류되는 것보다 자동차의 범주에 속하면서 제조회사는 ㈜기아자동차, 모델명은 K5, 제조년도는 2005년 등 보다 상세한 정보를 얻을 수 있다면 앞으로 다양한 응용 시스템에 활용할 수 있다.

본 논문에서는 딥러닝 기법에 의해 먼저 물체를 인식하여 물체에 대한 대분류 정보를 얻고, 이 정보를 인터넷 검색 툴의 쿼리 정보로 활용하여 보다 상세한 정보를 얻는 방법을 기술한다. 이전의 방법[1]과 비교하여 제안한 방법의 특징은 다음과 같다. 먼저 딥러닝 기법을 활용하여 물체인식을 수행하기 때문에 물체영역의 추출과 대분류에 대한 정확한 정보를 얻을 수 있다. 둘째, 그 결과를 인터넷 검색 툴의 쿼리 정보로 이용하여 인터넷으로부터 직접 2차적인 검색을 수행하기 때문에 새로운 물체에 대해 학습을 필요로 하지 않는다. 즉, 객체 인식으로부터 얻어지는 대분류 정보와 인터넷 검색으로부터 얻어지는 상세정보를 결합하여 검색에서 일반적으로 사용되는 coarse-fine procedure를 논리적으로 실현했다는 특징을 갖는다.

### 2. 관련 연구

#### 2.1 물체 검출

영상 속에 들어 있는 물체 검출은, 앞서 설명한 바와 같이, 대부분 2차원의 특징벡터에 근거한 기술자(descriptor)를 이용하여 모델과 영상의 매칭을 통해 이루어진다. 최근 들어, Deep Neural Network architecture가 물체인식에 도입되면서 R-CNN[2]을 시작으로 Faster R-CNN[3], YOLOv2[4] 등 많은 방법들이 제안되었다.

#### 2.2 YOLOv2 알고리즘

YOLO(You Only Look Once) v2는 2016년에 제안된 알고리즘으로, 영상으로부터 45프레임 정도의 실시간으로 물체를 검출할 수 있는 알고리즘이다. 이 YOLO v2는 이전 알고리즘 YOLO에 비해 성능이 빨라지면서 정확도도 한층 높아졌다. 중요한 특징으로는 Batch Normalization[5] 기법을 이용하여 Training Dataset 전체에 대해 평균과 분산값을 구하는 것이 아니라, mini-batch 단위로 계산하는 방법과 High-resolution Classifier를 이용했기 때문에 정확도를 높일 수 있다.

### 3. 물체검색 기법을 결합한 상세정보 검색

구현은 영상을 입력받고 YOLO v2 알고리즘을 이용해 범주와 위치 결과를 이용하여 부분 영상을 생성한다. 부분 이미지를 이용하여 각 사이트별로 영상 검색을 실시하고 이후 알고리즘을 이용해 세부정보를 취득한다.

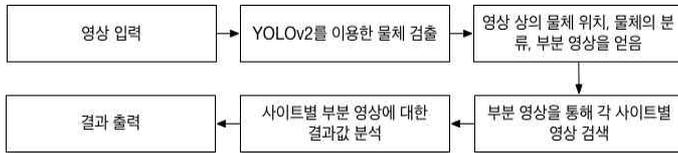


그림 1. 전체적인 알고리즘 흐름도

먼저 영상을 받아 딥러닝 기반 물체검출 방법을 이용해 물체를 선택하고(결과 값이 여러 개인 경우) 사이트별로 영상을 검색한다. 이전의 방법[1]은 YOLO를 사용한 반면, 제안한 방법은 YOLOv2를 사용하고 있다.

영상 검색은 사용자가 많은 사이트인 Google, Baidu, Yandex 사이트를 이용했고, 각각의 사이트에서 검색된 데이터를 이용하여 영상에 대한 데이터를 추출했다.

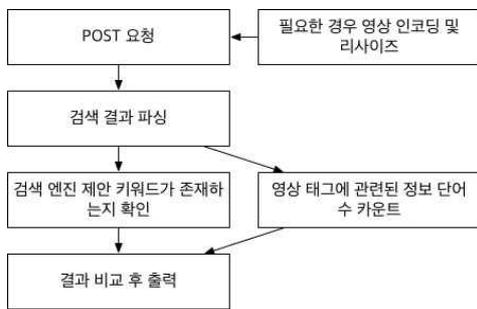


그림 2. 데이터 분석 알고리즘 흐름도

먼저 사용자가 검색할 물체를 선택하면 그에 해당되는 영상 파일을 각 사이트별로 검색하여 그에 대한 결과를 HTML 파싱을 통해 가져오게 된다.

파싱은 Google 사이트의 경우 이 영상에 대한 검색어, 검색된 영상과 유사한 영상으로부터 정보를 찾는다. 결과 값이 적거나 유사한 영상으로부터 정보를 찾지 못한다면 결과로는 관련 영상만 출력된다.

파싱된 결과 값의 분석은 쓸데없이 파싱된 결과를 제거하고, 영상에 대한 가장 빈도수가 많은 단어를 분석한다. 만약 결과로 출력된 값이 물체의 분류와 관련 없는 값일 경우도 있는데, 이 경우는 오류로 표기했다.

이전의 방법[1]과 제안한 방법은 데이터 분석 알고리즘의 검색엔진 제안 키워드와 관련된 단어를 카운트하여 두 가지 변수의 비중을 다르게 두었다는 점이 다르다.

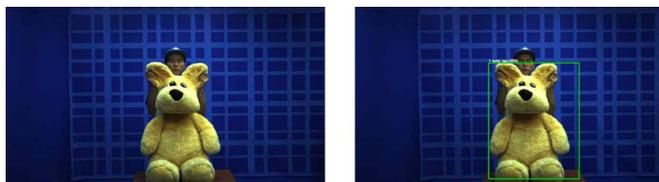


그림 3. 원본 영상 (왼쪽), 물체 검출의 결과(오른쪽)

Google	teddy bear
Baidu	泰迪熊
Yandex	teddy bear

표 1. 그림 3의 사이트별 검색 결과

결과 평가는 몇몇 영상을 선택하여 영상 객체에 대한 정보 이외에 추가적인 정보를 얻어진 경우와 그렇지 않은 경우를 비교하여 평가했다.

분류	정답	오류	정확도
Person	12	8	60%
Objects	13	7	65%
Vehicle	12	8	60%

표 2. 실험 결과

#### 4. 결론

이 논문에서는 딥러닝을 이용한 물체검출 방법을 이용해 영상에 대한 상세정보를 알 수 있는 방법을 제안했다. 물체에 대한 분류는 물체검출 방법을 통해 알 수 있지만, 물체에 대한 정보는 물체검출 방법을 통해 알 수 없기 때문에 영상 검색을 통해 물체에 대한 정보를 알아내는 방법을 이용했다.

영상 검색을 이용하기 때문에 영상 검색엔진의 데이터의 양에 대해 결과가 좌우되는 것으로 보인다. 영상에서 다른 특징적인 영역이 없으면 비슷한 영상을 검색하여 찾는 방법은 그다지 도움이 되지 않는다. 반면에 영상에 대한 특징점이 뚜렷하고 비슷한 영상이 검색 엔진에 많이 존재하다면 물체에 대해 구체적인 정보를 찾을 수 있다.

#### 감사의 글

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 '범부처 Giga KOREA 사업'의 지원을 받아 수행된 연구임 (GK17C0100, 기가급 대용량 양방향 실감 콘텐츠 기술 개발)

#### 참고문헌

- [1] 김선우, 김연우, 이철우, “딥러닝을 이용한 객체 검출을 활용하여 이미지에 대한 정보 찾기 방법 제안”, 스마트미디어학회 춘계학술대회, 2017.
- [2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014.
- [3] S. Ren, K. He, R. Girshick, and J. Sun. “Faster r-cnn: Towards real-time object detection with region proposal networks”. arXiv preprint arXiv:1506.01497, 2015.
- [4] J. Redmon and A. Farhadi. “Yolo9000: Better, faster, stronger.” arXiv preprint arXiv:1612.08242, 2016.
- [5] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. ICML, 2015.