

# 물체 지역화와 심층 나선형 인공 신경망을 이용한 이미지 검색

조재형<sup>o</sup>, 윤성의

KAIST 전산학부

dil122001@kaist.ac.kr, sungeui@gmail.com

## 요약

본 논문은 최근 이미지 분류 연구에서 우수한 성능을 보이고 있는 심층 나선형 인공 신경망을 이용한 이미지 검색 방법을 제안하고 있다. 심층 나선형 인공 신경망의 중요한 특징 중 하나는, 후반부 은닉 층에서 추출한 벡터를 이미지 기술자로 사용시 고차원 개념을 잘 표현할 수 있다는 점이다. 이 특징에서 착안하여, 물체 지역화 기술을 이용해 이미지에서 물체가 있는 영역들을 찾아낸 뒤 해당 영역들에서만 기술자를 추출하였다. 그 결과, 서로 재훈련을 하지 않은 상태에서 심층 나선형 인공 신경망을 이용한 최신 이미지 검색 기술보다 높은 top-4 점수를 UKB 데이터셋으로부터 얻을 수 있었다. 이러한 결과는 물체 영역으로 한정하여 기술자를 추출함으로써 심층 나선형 인공 신경망에서 추출하는 이미지 기술자의 장점을 더욱 부각시킨 점에 기인한 것으로, 실험을 통해 제안한 방식의 우수성을 입증하였다.

## 1. 서론

이미지 검색은 주어진 질의 이미지와 유사한 이미지들을 찾아내는 기술이다. 이 기술은 질의 이미지가 무엇을 나타내고 있는지 판단하는 기술인 이미지 분류 기술과 일고자 하는 결과는 다르지만, 두 기술 모두 이미지에서 기술자를 추출하여 목적에 맞게 사용한다는 공통점이 있다. 기존에는 SIFT[13], GIST[17] 등의 사람의 통찰에 기반해 추출하는 기술자를 사용하여 이미지 검색, 분류를 하는 연구들이 많이 이루어져 왔다 [14, 15, 16]. 더불어 최근에는 심층 나선형 인공 신경망(CNNs) [5]을 이용한 이미지 분류 기술들이 활발하게 연구되고 있으며, 기존 기술들보다 높은 분류 정확도를 보이고 있다. CNNs를 분류기로서 사용한 연구뿐만 아니라, CNNs의 은닉 층에서 추출한 벡터를 이미지 기술자로서 사용한 연구들도 높은 성능을 보이고 있다 [1].

[1]가 높은 분류 정확도를 보인다는 것은 CNNs에서 추출한 기술자가 이미지의 특징을 잘 나타낸다는 것을 의미한다. 따라서 CNNs에서 추출한 기술자를 이미지 검색 기술에 적용해 보았을 때도 높은 성능을 보인다는 것이 최근 연구에 의해 알려져 있다 [2].

[2]에 따르면 CNNs의 7번째 은닉 층에서 추출한 기술자는 이미지의 질감과 같은 저차원의 정보보다는 이미지가 나타내고 있는 개념과 같은 고차원의 정보를 더 잘 나타낸다. 하지만 그림 1에서 볼 수 있듯이 이미지가 나타내는 개념은 이미지 전체에서 기술자를 추출할 때 보다, 이미지상에서 의미 있는 영역으로 범위를 한정하여 추출할 때 더욱 명확해진다. 따라서 이 논문에서는 이미지에서 물체가 있

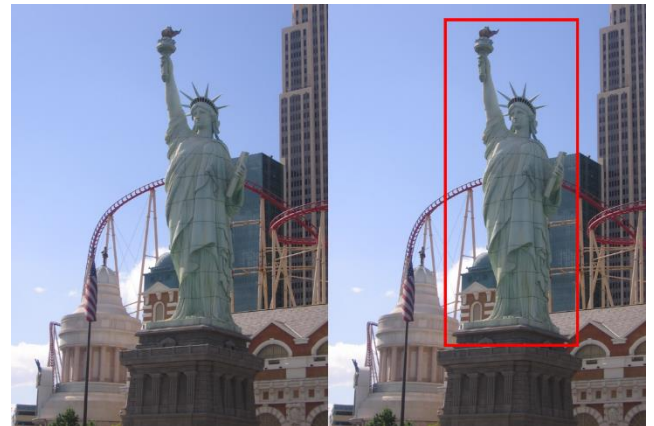


그림 1. 전체 이미지보다 물체 영역을 한정 지었을 때 나타내고자 하는 의미가 더 명확해진다.

는 영역을 찾아내는 물체 지역화 기술과 CNNs를 사용한 이미지 기술자 추출 방법을 이용하여 이미지를 검색하는 방법을 제안하고자 한다.

## 2. 관련 연구

이 논문은 이미지에서 물체를 검출하는 문제를 다룬 논문인 [1]에서 많은 영감을 얻었다. [1]에서는 selective search [11] 방식을 이용하여 물체가 있을 법한 후보 영역을 검출하고, 각각의 후보 영역들에 대하여 CNNs를 사용하여 이미지 기술자를 추출하였다. 추출된 기술자와 클래스마다 훈련되어 있는 이진 서포트 벡터 머신을 이용하여 해당 후보 영역이 어떤 물체를 나타내고 있는지 판별한다. 이 방식이 높은 성능을 보일 수 있는 가장 큰 이유는 전체 이

미지에 대한 기술자를 추출하는 대신에 영역을 물체 주위로 한정시켜 추출하여서, 기술자가 더 명확하게 물체를 나타내기 때문이다. 이러한 장점을 이미지 검색에도 적용하기 위하여, 본 논문에서는 [1]과 비슷하게 먼저 후보 영역을 검출한 뒤 각 후보 영역에 대해 이미지 기술자를 추출하였다.

하지만 이미지 검색을 위한 후보 영역 검출에는 selective search 방식이 적절하지 않다. Selective search는 이미지 분할을 이용해 계층적으로 후보 영역을 제안하는 방식으로, 여러 계층에 걸쳐서 최대한 많은 후보 영역을 제안한다. 그렇기 때문에 이미지에 포함된 물체를 검출하려 할 때는 효율적인 방식이다. 하지만 최대한 많은 후보 영역을 제안하는 것을 목표로 하다 보니 물체 영역의 수 보다 물체를 나타내지 않는, 의미 없는 영역의 수가 훨씬 많게 되는 현상을 초래한다. 이미지 검색을 위해 기술자를 추출할 때는, 이런 의미 없는 영역으로부터의 기술자는 오히려 검색 정확도를 낮추는 요인이 된다. 따라서 모든 후보 영역을 제안 받기 보다는 물체가 있을 법한, 의미 있는 영역만을 제안 받기 위해 물체도를 추정하여 후보 영역을 제안하는 방식을 사용하였다 [3]. 이 방식은 인간이 물체를 판별할 때, 대상이 어떤 물체인지 인식하기 전에 형태만으로 물체인지 아닌지를 먼저 판단하는 능력에서 영감을 얻었다. 이미지의 해상도를 변경해가며 영역들로부터 norm of gradient 를 계산해 물체도를 추정하는 방식인데, 이 계산을 이진화하여 더욱 가속화 시킴으로써 보다 효율적인 물체 영역 검출 방식이 되었다.

CNNs 로부터 추출한 이미지 기술자가 이미지 분류에서 좋은 성능을 보이는 것은 기존 연구를 통해 검증된 바 있다 [1]. 더불어 [2]를 통해, 이 기술자가 이미지 검색에도 유용하게 적용될 수 있음이 연구된 바 있다. [2]에서는 CNNs 의 5, 6, 7 번째 은닉 층으로부터 추출한 기술자들을 이용해 검색 정확도를 평가하였다. 그 결과, 5 번째 은닉 층에서 추출한 기술자는 이미지의 질감과 같은 저차원 정보를 잘 나타내고, 7 번째 은닉 층에서 추출한 기술자는 이미지가 나타내는 물체, 개념과 같은 고차원 정보를 잘 나타낸다는 것을 밝혔다. 또한 CNNs 에서 추출한 기술자는 주성분 분석을 통해 차원을 축소했을 때, 기존에 사용되던 기술자들에 비해 검색 정확도가 덜 하락한다는 것을 밝혔다.

### 3. 물체 지역화와 심층 나선형 인공 신경망을 이용한 이미지 검색 방법

이 장에서는 먼저 물체 지역화와 CNNs 를 이용해 이미지로부터 기술자를 추출하는 방법에 대해 설명하고, 추출된 기술자를 이용해 이미지들 간의 유사도를 측정하기 위한 거리 측정 방법을 설명한다.

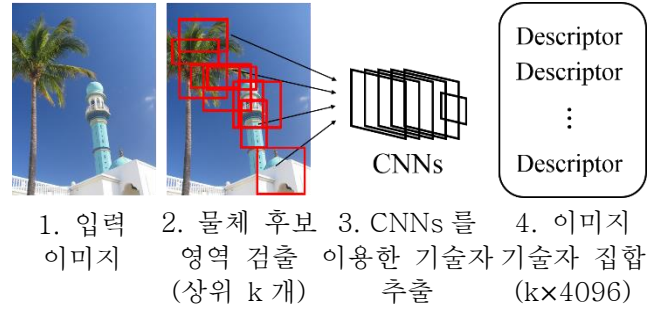


그림 2. 입력 이미지로부터 이미지 기술자 집합을 추출하는 과정

#### 3.1 기술자 추출 방법

기술자 추출 과정은 크게 물체 지역화 기술을 이용하여 물체 후보 영역을 검출하는 단계와 CNNs 를 이용하여 물체 후보 영역들로부터 기술자를 추출하는 단계로 나눌 수 있다. 먼저 물체 후보 영역 검출 단계에서는 [3]에서 제안한 BING 물체도 예측 방식을 사용하였다. 2 장에서 설명한 바와 같이 BING 물체도 예측 방식은 각 후보 영역들에 대하여 물체도를 수치적으로 추정하기 때문에, 물체도가 높은 영역들만 선택적으로 사용할 수 있다. 또한 물체도 예측에 필요한 연산을 이진화하여 후보 영역 제안 속도를 가속화시켰기 때문에, [1]에서 후보 영역 검출에 사용한 selective search 방식에 비하여 검출 속도가 훨씬 빠르다. 이러한 장점들 덕분에, 기술자의 크기와 처리 속도가 중요한 이미지 검색에는 BING 물체도 예측 방식이 더욱 적합하므로 본 연구에서는 이 방식을 이용하여 후보 영역을 추출하였다.

BING 물체도 예측 방식을 사용하면 많은 물체 후보 영역을 검출할 수 있고, 각각의 후보 영역에 대하여 수치적인 물체도 값을 얻을 수 있다. 하지만 물체 후보 영역 중에는 실제로 물체를 나타내는 영역 보다 이미지를 대표할 수 없는, 의미 없는 영역의 비율이 훨씬 크다. 따라서 이미지를 대표할 수 있는 물체 영역들에서만 기술자를 추출하기 위하여 물체도 값을 기준으로 영역들을 정렬시킨 뒤, 물체도 값이 가장 큰 k 개의 영역들에서만 기술자를 추출하였다.

영역으로부터의 이미지 기술자 추출은 Caffe 에서 제공하는 모델인 BVLC Reference CaffeNet 을 사용하였다 [4]. 이 모델은 AlexNet[5]을 ILSVRC 2012[6]로 학습시킨 CNNs 모델이다. 이 모델의 출력 층은 1000 차원 벡터인데 본 논문에서는 [1], [2]처럼 출력 층 이전의 은닉 층들로부터 얻을 수 있는 4096 차원 벡터를 이미지 기술자로서 사용하였다. 그림 2 에는 이미지로부터 이미지 기술자 집합을 추출하는 과정이 요약 되어있다.

#### 3.2 이미지 간의 거리 측정 방법

위에서 설명한 방법으로 이미지 기술자를 추출



그림 3. 이미지간 거리 계산 방법의 시각화  
왼쪽 이미지와 오른쪽 이미지의 물체 영역들 중, 빨간색으로 표시한 물체 영역간의 L1 거리가 가장 가깝다. 따라서 이 물체 영역간의 거리가 두 이미지간의 거리를 나타내게 된다.

하면 각각의 이미지는 k 개의 4096 차원 벡터로 나타내어지게 된다. 이미지들간의 유사도를 측정하기 위해서는 이미지 기술자들 간의 거리를 측정할 수 있어야 한다. 따라서 본 논문에서는 이미지 기술자들의 집합 X, Y 간의 거리를 다음과 같이 정의하였다.

$$\text{dist}(X, Y) = \min(d(x, y)) \quad \forall x \in X \quad \forall y \in Y$$

여기서  $d(x, y)$ 는 두 벡터간의 거리를 나타내고, 본 논문에서는 여러 거리 계산법 중 L1 거리 계산법을 사용하였다. 각각의 이미지 기술자들은 전체 이미지 안에서 물체 영역을 나타내고 있다. 따라서 그림 3 에서 볼 수 있듯이 위의 거리 정의는 두 이미지간의 거리를 각각이 포함하고 있는 물체들 중 가장 비슷한 물체들 간의 거리로 나타내겠다는 것을 의미한다.

또한 본 논문에서는 위의 거리 계산 방식을 투표 방식에도 적용시켜보았다. 투표 방식 중, 거리가 가장 가까운 한 후보에만 투표하는 대신 여러 후보에게 가우시안 확률 값에 따라 투표 점수를 다르게 부여하는 soft voting [8] 방식에서 영감을 얻어, 각각의 물체 영역들로부터 가장 가까운 k 개의 물체 영역들을 찾고, 찾아진 물체 영역들이 포함된 이미지에 순위에 맞는 투표 점수를 부여하였다. 즉, 한 번의 질의에 총  $k \times k$  개의 투표가 이루어지고, 각각의 투표는 질의 물체 영역과 가까운 순위에 따라 점수를 다르게 부여 받게 된다.

#### 4. 실험 결과 및 분석

3 장에서 설명했듯이 이 방법은 BING 물체도 예측 방식과 CNNs 를 이용하여 이미지 기술자를 추출한다. 따라서 Python 에서 구현된 BING 물체도 예측 방식 코드[3, 7]와 Caffe 에서 제공되는 R-CNN[1]에 기반한 물체 검출 코드를 구현에 이용하였다.

실험에는 INRIA Holidays dataset 을 사용하였다 [9]. 이 데이터셋은 500 여 가지의 물체와 장소를 찍은

사진들로 구성되어 있으며 총 1491 장의 사진을 담고 있다. 평가는 실험 이미지들을 질의 이미지와의 유사도를 기준으로 정렬했을 때, 정답 이미지들의 순위 값에 의해 계산되는 mAP 점수를 사용하여 이루어진다. 그리고 최종 평가에는 INRIA Holiday dataset 과 UKB dataset [10]을 사용하였다. UKB 데이터셋은 2550 가지의 물체를 실내에서 찍은 10200 장의 사진들로 구성되어있으며, 평가는 질의 이미지와 가장 가까운 4 장의 이미지를 찾아, 그 안에 정답 이미지가 얼마나 포함되어 있는지 판단하는 top-4 점수를 사용한다.

제안된 이미지 기술자 추출 방법에는 CNNs 로부터 이미지 기술자를 추출할 때 사용하는 은닉 층의 선택과 이미지를 기술자의 집합으로 나타낼 때 사용하는 물체 영역의 개수 k, 이 두 가지 변수가 존재한다. 따라서, 이 장에서는 먼저 사용하는 은닉 층에 따른 성능 차이를 분석하고 물체 영역의 개수 k 값에 따른 성능의 변화를 분석 한다. 그리고 마지막으로 현재 본 논문에서 제안하는 방식의 성능과 다른 최신 방식들의 성능을 비교분석 한다.

#### 4.1 이미지 기술자를 추출하는 은닉 층에 따른 성능 분석

[2]에서도 CNNs 를 이용해 이미지 기술자를 추출하였으며, 이 때 기술자를 추출하는 은닉 층에 따라서 성능이 바뀌었다. 따라서 본 논문에서도 어떤 은닉 층이 본 논문에서 제안하는 기술자 추출 방식에 적합한지 알아보기 위해 실험을 해 보았다. 물체 영역의 개수 k 는 10 으로 고정하였고 투표방식은 사용하지 않고 거리 비교 만으로 이미지 검색 성능을 측정하였다. 5 번째(pool5), 6 번째(fc6), 그리고 7 번째(fc7) 은닉 층을 비교해 보았고, 결과는 다음과 같다.

표 1. 은닉 층에 따른 검색 성능 비교 (INRIA Holidays dataset)

은닉 층	pool5	fc6	fc7
mAP 점수	0.25247	0.49352	0.56119

위의 표 1에서 볼 수 있듯이 후반부의 은닉 층으로 갈수록 검색 정확도가 올라가는 것을 볼 수 있다. [2]에 따르면 5 번째 은닉 층은 이미지의 질감과 같은 저차원 정보를 잘 나타내고, 7 번째 은닉 층은 이미지가 나타내는 물체, 개념과 같은 고차원 정보를 잘 나타낸다. 또한 [2]의 성능 분석에서는 6 번째 은닉 층의 성능이 가장 높았고, 5 번째 은닉 층의 성능이 본 논문의 결과처럼 눈에 띄게 낮지 않았다. [2]와 본 논문의 방식이 이런 추세 차이를 보이는 이유는, 물체 지역화 기술의 사용 때문이다. [2]에서는 전체 이미지에서 기술자를 추출하기 때문에 풍경 사진에서는 저차원의 질감 정보를, 물체 사진에서는 고차원의 물체 정보를 잘 나타낼 필요가 있다. 따라서 이 두 정보를 적절히



담고 있는 6 번째 은닉 층이 가장 높은 성능을 보인다. 하지만 본 논문이 제안하는 방식에서는 물체 지역화 기술을 이용하여 기술자를 추출하는 영역을 물체 영역으로 한정한다. 즉, 5 번째 은닉 층은 저차원의 질감 정보를 나타내기 때문에 물체 영역에서 해당 물체들간의 고차원 정보의 차이를 잘 구별하지 못한다. 따라서 본 논문이 제안하는 기술자 추출 방법에서는 물체 영역 안에서 고차원의 특징 정보를 잘 구별하는 7 번째 은닉 층이 적합하다는 것을 알 수 있다. 이러한 이유로 이후의 실험에서는 모두 7 번째 은닉 층에서 기술자를 추출하여 사용하였다.

### 4.2 이미지를 나타내는 물체 영역의 개수에 따른 성능 분석

다음으로는 이미지를 나타내기 위해 기술자를 추출하는 물체 영역의 수  $k$ 에 따른 이미지 검색 성능을 분석한다. 이 실험에서도 투표 방식을 사용하지 않고 거리 비교만으로 검색 성능을 측정하였다. 위에서 설명했듯이 기술자는 7 번째 은닉 층에서 추출하였다.  $k$  값은 5, 10, 15, 25, 50 으로 바꾸면서 실험하였고, 그 결과는 다음과 같다.

표 2.  $k$  값에 따른 검색 성능 비교 (INRIA Holidays dataset)

k	5	10	15	25	50
mAP	0.5226	0.5612	0.5933	0.6150	0.6172

표 2 를 통해 알 수 있듯이 사용하는 물체 영역의 수가 많아질수록 mAP 점수가 증가한다. 하지만  $k$  값이 커질수록 증가하는 mAP 점수의 폭은 점점 줄어든다. 그 이유는 물체 영역들을 물체도 값이 높은 영역부터 순서대로 사용하기 때문에 의미 있는 후보 영역들은 이미  $k$  값이 작을 때부터 이미지를 나타내기 위해 사용되고 있기 때문이다. 따라서  $k$  값이 커져서 추가되는 영역이 생기더라도 그 영역들이 새로운 의미 있는 물체를 나타내는 경우가 점점 드물어 진다. 그러므로  $k$  값이 커짐에 따라 mAP 점수의 증가 폭이 줄어들게 되는 것이다. 또한  $k$  값이 크다는 것은 하나의 이미지를 나타내기 위해 사용되는 이미지 기술자의 개수가 많다는 것이므로, 하나의 이미지를 나타내는 기술자 집합의 크기가 크다는 것을 의미한다. 따라서 효율적인 이미지 검색을 위해서는 mAP 점수만을 보고 높은  $k$  값을 사용하기 보다는, 목표 데이터셋의 크기와 사용하는 시스템의 성능을 고려하여 검색 정확도와 기술자 집합의 크기를 절충하는  $k$  값을 선택하는 것이 중요하다.

### 4.3 최신 기술들과의 성능 비교 분석

마지막으로 본 논문에서 제안하는 방식과 다른

최신 이미지 검색 기술들의 성능을 비교해 보았다. 성능 비교에는 INRIA Holidays dataset 을 이용한 mAP 점수와 UKB dataset 을 이용한 top-4 점수를 사용하였다. 현재 UKB dataset 에서 가장 좋은 성능을 보이는 것으로 알려진 Sparse-coded features[12]와 INRIA Holidays dataset 에서 가장 좋은 성능을 보이는 것으로 알려진 Neural codes[2]와 비교를 해 보았다. 추가적으로 Neural codes 기술은 본 논문과 비슷하게 CNNs 를 이용하여 추출한 이미지 기술자를 사용한 이미지 검색 방식이므로 더 자세히 성능 비교를 하였다.

표 3. 최신 검색 기술들과의 성능 비교

기술자	UKB	Holidays
Ours(without voting)	3.28	0.617
Ours(with voting)	3.53	0.738
Neural codes(without retraining)[2]	3.43	0.749**
Neural codes(retraining on turntable views)[2]	3.56*	0.754**
Neural codes(retraining on Landmarks dataset)[2]	3.29	0.793*,**
Sparsed-coded features[12]	3.76	0.767

\*가 표시된 값은 CNNs 를 목표 데이터셋에 의존적으로 재훈련하여 얻은 결과임.

\*\*가 표시된 값은 Holidays dataset 에서 정방향이나 인 사진들을 수작업으로 정방향으로 회전시킨, 수정된 dataset 에서 얻은 결과임.

표 3에서 볼 수 있듯이 본 논문의 방식은 투표 방식을 적용함으로써 검색 정확도를 향상시켰지만 현재 최신 기술들의 성능을 넘지는 못했다. 하지만 본 논문의 방식처럼 CNNs 로부터 추출한 기술자를 사용하여 최고의 성능을 보이고 있는 Neural codes 방식과 조금 더 자세히 비교를 해 보고자 한다. 먼저 Holidays 에서 가장 높은 성능을 보이는 Neural codes 방식은 기술자 추출에 사용하는 CNNs 를 Holidays 데이터셋에 보다 적합하도록 fine-tuning 하기 위해 Landmarks 데이터셋으로 CNNs 를 재훈련하였다. 그 결과 Holidays 에서의 mAP 점수는 크게 증가하였지만, 오히려 UKB 에서의 top-4 점수는 감소하였다. 또한 UKB 에 적합하도록 CNNs 를 재훈련 한 경우에도 Holidays 에서의 점수는 큰 변화가 없지만 UKB 에서의 top-4 점수는 크게 증가한 것을 볼 수 있다. 여기서 알 수 있듯이 CNNs 를 목표로 하는 데이터셋과 유사한 데이터셋으로 재훈련 할 경우, 위에서 UKB 의 top-4 점수가 감소하는 경우처럼 기술자의 일반성이 떨어지거나 과잉적합이 발생할 수 있지만, 평가를 진행하는 대상인 목표 데이터셋에서 만큼은 큰 폭의 성능 향상을 보여준다. CNNs 의 재훈련이 성능에 큰 영향을 미치기 때문에, 목표 데이터셋에 맞춰 재훈련을 하지 않은 현재의 본 논문의 방식으로는 각각의 데이터셋에 맞게 재훈련 된 Neural codes 의 성능을 뛰어넘지는 못했다. 하지만 서로 재훈련이 되지 않은 상태에서의

성능을 비교하면 투표 방식을 사용한 본 논문의 방식이 UKB 데이터셋에서 Neural codes 방식 보다 더 좋은 성능을 보인다. 오히려 본 논문에서 제안하는 방식이 UKB 에 맞춰 재훈련 된 Neural codes 에 더 가까운 성능을 보인다. 또한 서로 재훈련 되지 않은 상태에서 Holidays 데이터셋의 점수는 본 논문의 방식이 조금 낮게 측정되지만, 사실 Neural codes 방식의 mAP 점수는 Holidays 데이터셋에서 성능을 측정할 때 데이터셋에 포함된 회전된 사진들을 임의로 모두 정방향으로 만든 뒤 측정한 결과이다. 이 점을 고려한다면 서로 재훈련을 하지 않은 상태에서의 Holidays 데이터셋에 대한 성능은 비슷하다고 할 수 있다. 결과적으로, 재훈련 전의 상태를 비교한다면 본 논문에서 제안하는 방식이 Neural codes 방식보다 좋은 성능을 보인다.

## 5. 결론

본 논문에서는 물체 지역화 기술을 이용하여 심층 나선형 인공 신경망으로부터 추출하는 기술자의 장점을 부각시킴으로써 검색 정확도를 향상시킬 수 있는 방식을 제안하였다. 본 논문에서 제안하는 방식처럼 심층 나선형 인공 신경망을 이용해 이미지 기술자를 추출하고, 현재 Holidays 데이터셋에서 최고의 성능을 보이는 Neural codes 방식과의 비교를 통해 제안하는 기술의 성능을 평가하였다. 아직까지는 인공 신경망을 목적에 맞게 재훈련 하지 못했기 때문에 최종적인 성능 비교는 하지 못하였다. 하지만 서로 재훈련 하지 않은 상태를 비교하였을 때, 투표 방식을 적용한 본 논문의 방식이 Neural codes 방식 보다 더 좋은 성능을 보인다. 또한 UKB 데이터셋에서의 성능은 본 논문의 방식이 이미 재훈련을 한 Neural codes 방식과 가까운 성능을 보이고 있다. 현재 많은 심층 나선형 인공 신경망을 사용하는 기술들이 재훈련을 통해 큰 폭의 성능 향상을 보이고 있으므로, 본 논문에서 제안하는 방식도 목적에 맞는 데이터셋을 찾아 인공 신경망을 재훈련 한다면 지금보다 더욱 높은 검색 정확도를 얻을 수 있을 것이다. 또한 본 논문에서 제안하는 방식의 검색 정확도는 물체 영역 검출 성능에 크게 의존적이므로, 현재보다 고성능의 물체 지역화 기술이 개발된다면 본 논문에서 제안한 방식의 검색 정확도 또한 더욱 높아질 것이다.

## 감사의 글

리뷰를 도와주신 김동혁, 윤용직, 이윤석님께 감사드립니다.

본 연구는 미래창조과학부 및 정보통신기술연구원 홍센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [ R0126-15-1108 , 대용량 이미지 검색과 시제품 렌더링을 위한 근접질의 SW 개발]

이 논문은 2015 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2011-0030079)

## 참고문헌

- [1] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
- [2] A. Babenko, A. Slesarev, A. Chigorin, V. Lempitsky, "Neural Codes for Image Retrieval", European Conference on Computer Vision (ECCV), 2014.
- [3] M. Cheng, Z. Zhang, W. Lin, P. Torr, "BING: Binarized Normed Gradients for Objectness Estimation at 300fps", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
- [4] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. <http://caffe.berkeleyvision.org/>, 2013.
- [5] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks", NIPS, 2012.
- [6] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012). <http://www.image-net.org/challenges/LSVRC/2012/>.
- [7] A. Ferrari, BING-Objectness, <https://github.com/alessandroferrari/BING-Objectness>, 2015.
- [8] H. B. Mitchell, P. A. Schaefer, "A "Soft" K-Nearest Neighbor Voting Scheme", International Journal of Intelligent Systems, Volume 16, Issue 4, APR 2001.
- [9] Jegou, H., Douze, M., INRIA Holidays dataset, <http://lear.inrialpes.fr/people/jegou/data.php>, 2008.
- [10] Nist'er, D., Stew'enius, H., "Scalable recognition with a vocabulary tree", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.
- [11] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, "Selective search for object recognition", IJCV, 2013.
- [12] Ge, T., Ke, Q., Sun, J., "Sparse-coded features for image retrieval", British Machine Vision Conference, 2013.
- [13] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", IJCV, 2004.
- [14] S. Lazebnik, C. Schmid, J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.
- [15] Y. Zhang, Z. Jia, T. Chen, "Image retrieval with geometry-preserving visual phrases", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011.
- [16] X. Shen, Z. Lin, J. Brandt, S. Avidan, Y. Wu, "Object retrieval and localization with spatially-constrained similarity measure and k-NN re-ranking", IEEE Conference on Computer Vision and Pattern

Recognition (CVPR), 2012.

- [17] A. Oliva, A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope", International Journal of Computer Vision, Vol. 42(3): 145-175, 2001.