

CNN을 이용한 소셜 이미지 자동 태깅 **(Automatic Tagging for Social Images using** **Convolution Neural Networks)**

장 현 웅 (Hyunwoong Jang)
조 수 선 (Soosun Cho)

2016-01
[KCI 우수 등재]

1. 서론

[그림 1]과 같이 최근 인스타그램과 같은 소셜 미디어 공유 사이트에서는 이미지 검색을 이미지에 달린 태그를 기반으로 하였다.

이 방법의 문제점은, 태그는 각 개인의 주관적인 판단이 들어가서 원래 이미지의 정보와는 다른 '잡음태그'를 포함하고 있다는 점이다.

이러한 잡음태그는 정확한 검색에 방해가 되므로, 이미지에 대한 정확한 태깅의 필요성이 부각되고 있다.

하여, 기존의 **BoVW(Bag of Visual Words)**을 기반으로 한 이미지를 분류하던 방법과 비교하여 딥러닝 기반의 CNN을 사용하여 자동으로 이미지에 태그를 붙이는 **이미지 자동태깅** 방법에 대해 소개한다.



[그림1] 인스타그램의 태그시스템

2. 관련 연구

- BoVW(Bag of Visual Words) – 이미지 단어 분류

BoVW는 단어 모음(BOW)개념에서 채택된 방법이다. BOW는 어떠한 문서에서 각 **단어의 수**를 문서에 나타내어 해당 문서의 키워드를 알아내는 방법이다. 이와 같은 방법에서, **단어 대신에 이미지**를 "단어"로 사용한 방법이다. [그림2]와 같이 특징이 되는 이미지들을 모은 가방이라고 생각하면 된다.

- SIFT 알고리즘 – BoVW의 기반이 되는 알고리즘

기본적으로 미분을 이용하여 이미지 **화소 값의 변화 정도**를 측정하여 변화가 큰 영역들을 뽑아내어 '특징 벡터'를 구성하는 방법.
뚜렷한 경계를 지닌 객체의 이미지에서 효과적인 방법이나, 경계가 뚜렷하지 않는 배경 이미지와 같은 객체에서는 별로 효과를 보지 못하는 방법이다.



[그림2] 이미지 조각 가방

2. 관련 연구

- 딥러닝 기반 CNN(Convolutional Neural Network)

[그림3]은 '공간필터'라는 개념을 설명한 그림이다. [그림3]에서는

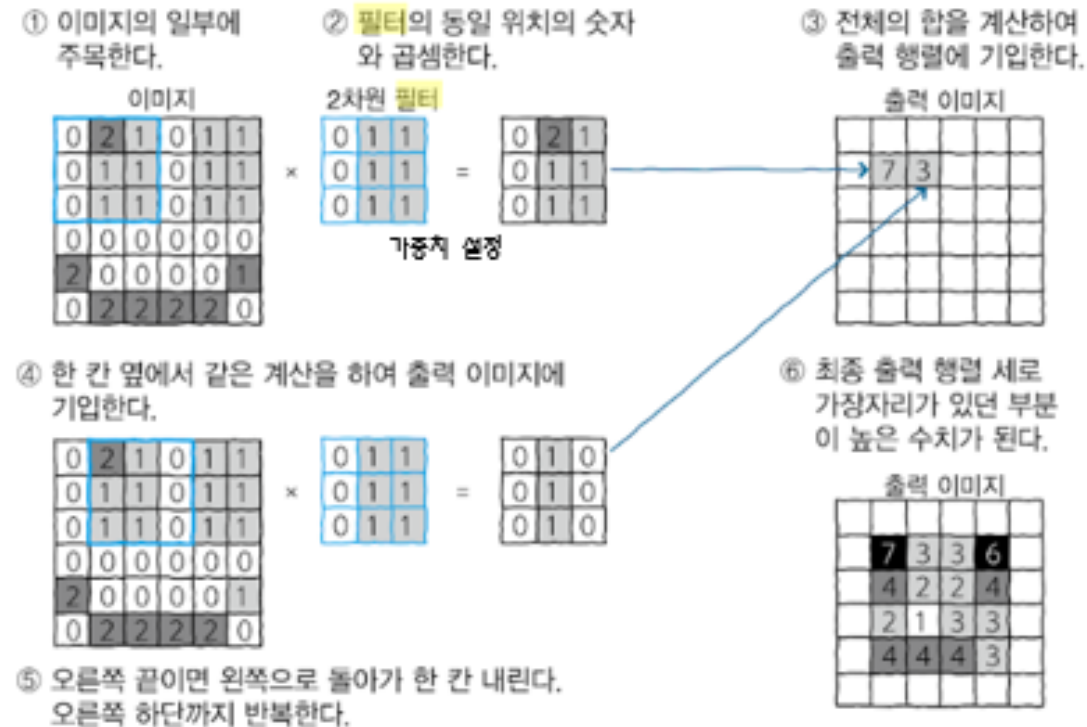
$$\begin{matrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{matrix}$$
 을 필터로 사용한 예이다. 어떠한 필터를 사용하느냐에

따라 이미지에서 강조되는 부분이 달라지므로, 어떠한 필터를

적용하냐도 굉장히 중요한 문제다. [그림4]를 예로 보면,

세로라인 왼쪽 강조 필터를 적용시, 세로라인 왼쪽 영역이

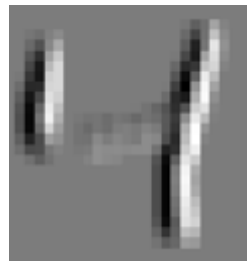
짙은 검정색으로 강조되는 것을 볼 수 있다.



[그림3] 공간필터



$$\begin{matrix} -2 & 1 & 1 \\ -2 & 1 & 1 \\ -2 & 1 & 1 \end{matrix}$$



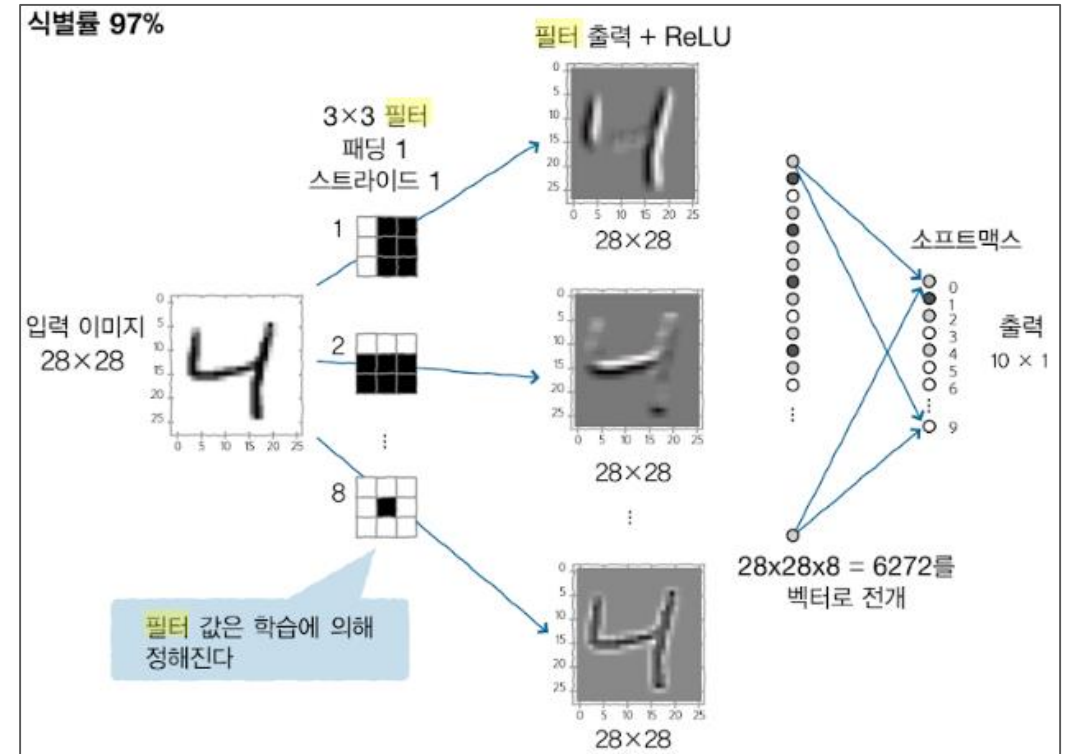
[그림4] 세로라인 왼쪽 강조 필터 적용시

2. 관련 연구

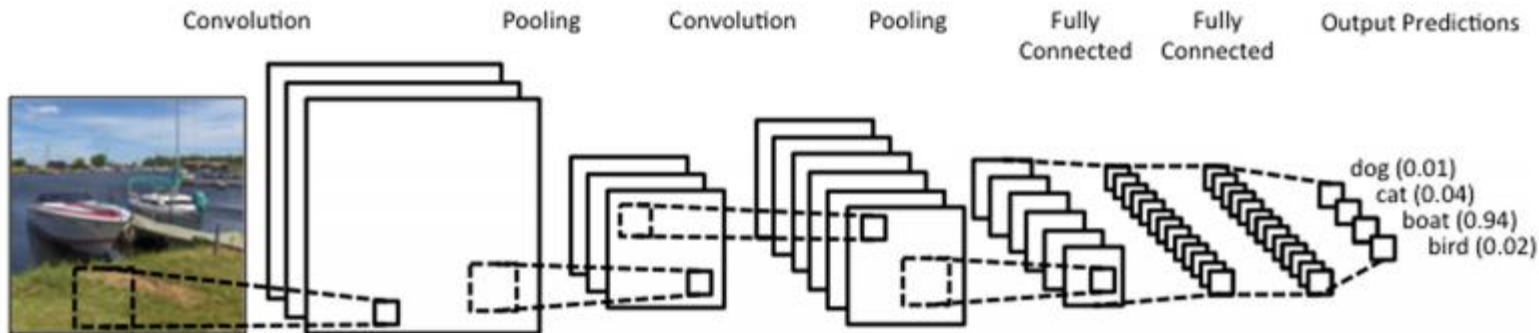
- 딥러닝 기반 CNN(Convolutional Neural Network)

이러한 필터를 딥러닝으로 학습시켜, **최적의 필터**를 찾고 이 필터를 이용하여 이미지를 분류하는 방법이 CNN이다. ([그림5])
이러한 개념을 바탕으로 [그림6]와 같은 층을 만들어서 학습시킨다.
이러한 방법으로, 별도의 특징 추출 알고리즘 없이 패턴을 인식하고 분류할 수 있다.

이 방법으로는 앞의 BoVW방법에서 약했던 배경이미지 인식에서도 좋은 결과를 얻을 수 있다.



[그림5] CNN



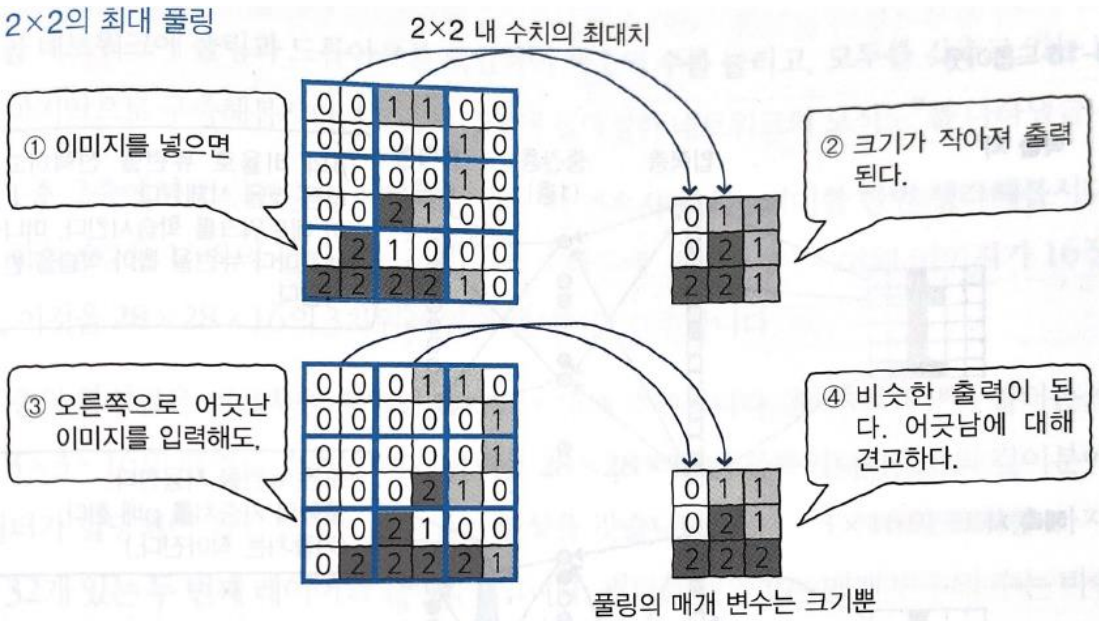
[그림6]

2. 관련 연구

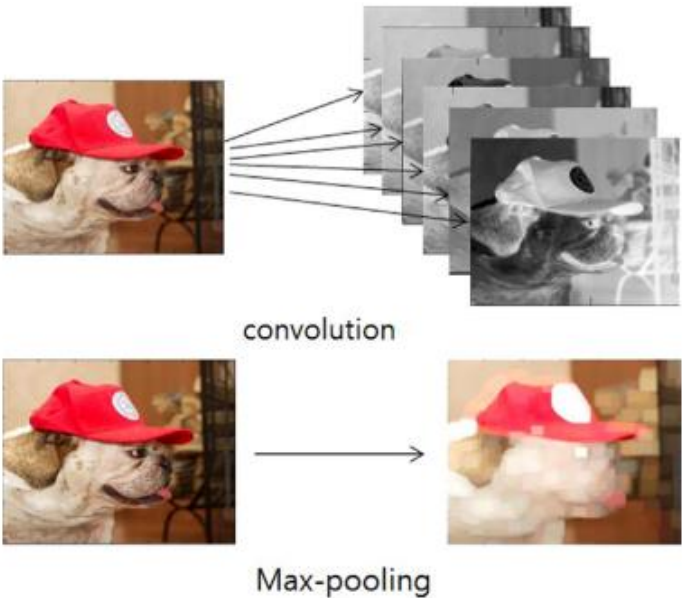
- 풀링(Pooling)

이미지 인식과 관련해 중요한 것이 이미지의 위치이다. 예를 들어, 필기체 숫자 '2'가 몇 픽셀만 어긋나도 각 배열의 수치가 완전히 달라진다. 이러한 문제를 해결하는 방법이 '풀링 처리'이다.

[그림7]에서와 같이 입력이미지가 가로 세로로 어긋나게되도, 비슷한 출력을 얻을 수 있게되는 성질을 부여할 수 있다.



[그림7]



[그림8]

3. 구현 및 실험

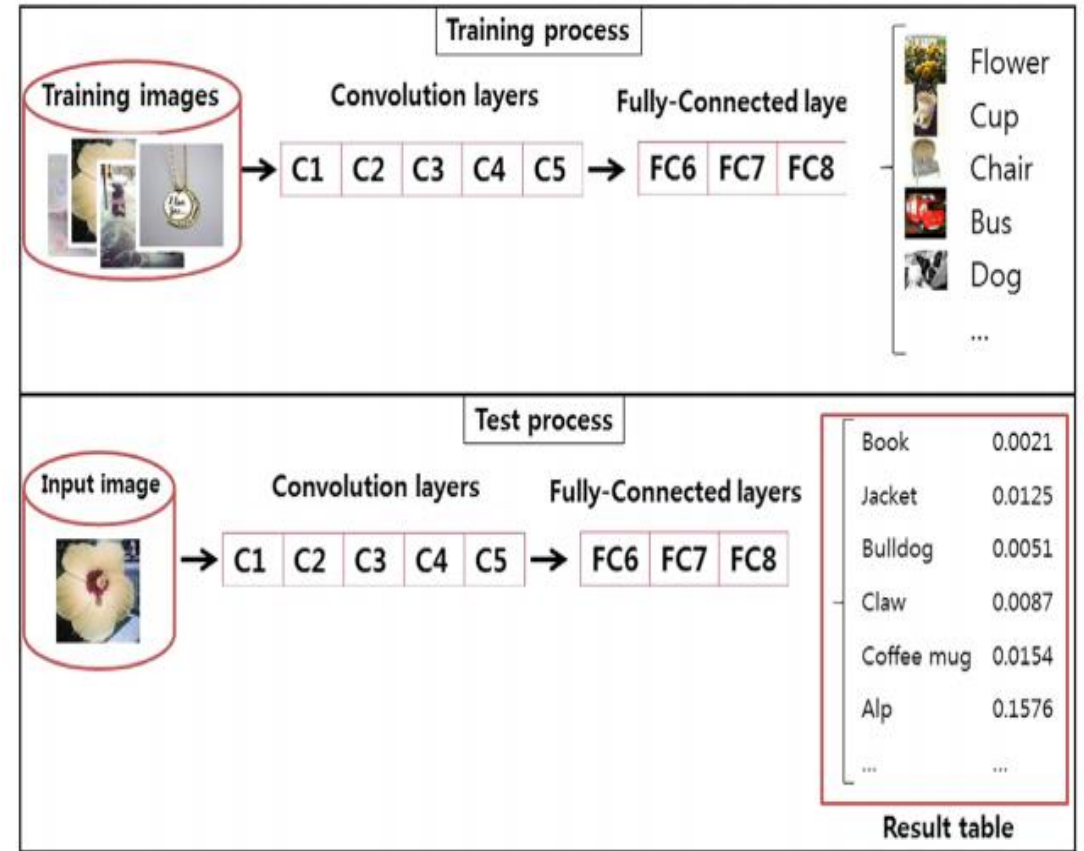
앞의 내용들을 바탕으로, 토론토대학의 Network모델을 사용하였고, 인스타그램의 이미지를 테스트 데이터셋으로 사용하였다.

전체적으로 5개의 convolution, max-pooling, normalization 계층과 3개의 fully-connected 계층으로 구성했다.

활성화 함수는 ReLU(Rectified Linear Unit)함수를 사용했다. 기존의 sigmoid함수가 아닌 ReLU를 사용하면서 계산량은 줄고 정확도는 증가했다.







ReLU는 0이상의 값은 출력해주고, 0이하인 값들을 0으로 만들어 계산량을 줄일 수 있었다.

정확하고 의미 있는 특징을 추출하기 위해서 weight인 convolution 필터를 훈련시키는데, BackPropagation과 경사하강법을 통해 이루어진다. [그림9]의 FC6과 FC7 계층에서 지속적인 훈련과 특징을 추출하고, 마지막 계층에서 softmax로 분류를 하게 된다.



[그림9]

4. 결과분석 및 평가

Image	Instagram Manual tags	Automatic tags using CNN	Image	Instagram Manual tags	Automatic tags using CNN	Image	Instagram Manual tags	Automatic tags using CNN
	bike smoke pretty	chain necklace bottlecap		bird photography ocean sun	albatross mollymawk pelican goose		bottle nice style	stole wool prayer rug prayer mat
	car accordion travel	wing valley vale alp		chair dog design pet	miniature poodle standard poodle toy poodle		house office luxury terrace	patio terrace home theater monitor

위의 표는 인스타그램에서 검색한 이미지에 달린 태그들과 같은 이미지를 CNN을 통해 분석하고 태그를 붙인 결과를 비교한 것이다. CNN에 의한 자동 태깅 결과, 기존에 붙은 태그보다 더 섬세하고 정확한 태그가 붙여진 것을 볼 수 있다. 기존의 인스타그램 고유의 태그를 이용한 검색정확도인 60.89%에서 본 연구에서 제안한 방법의 결과인 70.44%로 약 10% 정확도가 증가한 것을 알 수 있다.