

CNN을 이용한 소셜 이미지 자동 태깅

Automatic Tagging for Social Images using Convolutional Neural Network

장현웅, 조수선 (2016). 출처 : 정보과학회논문지, 발행 : 한국정보과학회

KAIG 20190312

발표자 : 조소영

목차

0. 요약

1. 서론

- 기존의 이미지 분석 방법
- CNN 기반 이미지 분석 방법

2. 본론 (구현 및 실험)

3. 결론 (결과 분석)

0. 요약 (abstract)

연구 배경 : 소셜 미디어 공유 사이트 → 일반적으로 이미지 태그 정보 사용
But, 이미지 & 영상 데이터의 폭발적 증가

- 1) 이미지 태깅 번거로움 (수작업의 한계)
- 2) 잘못 붙여진 태그 / 태그가 없는 경우 → 이미지 검색 정확도 ↓

∴ **논문 주제** : 자동으로 이미지로부터 태그 추출하는 방법 제시

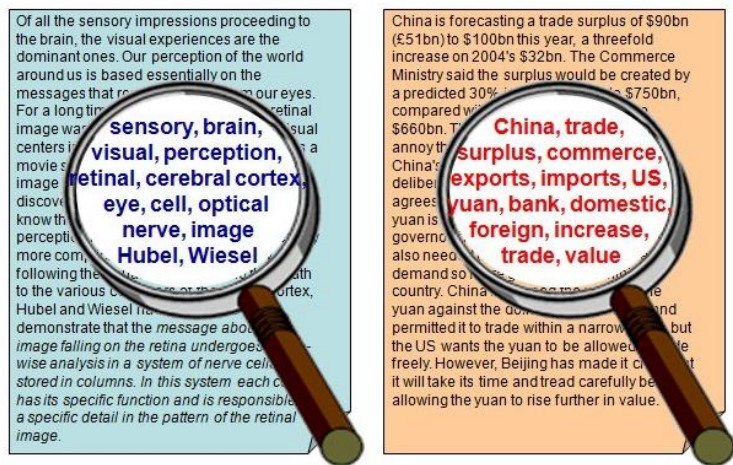
학습(train) : CNN 기법으로 ImageNet 제공 대용량 이미지 데이터&라벨 (태)

테스트(test) : 인스타그램 이미지에서 라벨 정보 추출, 자동 태깅

1. 서론

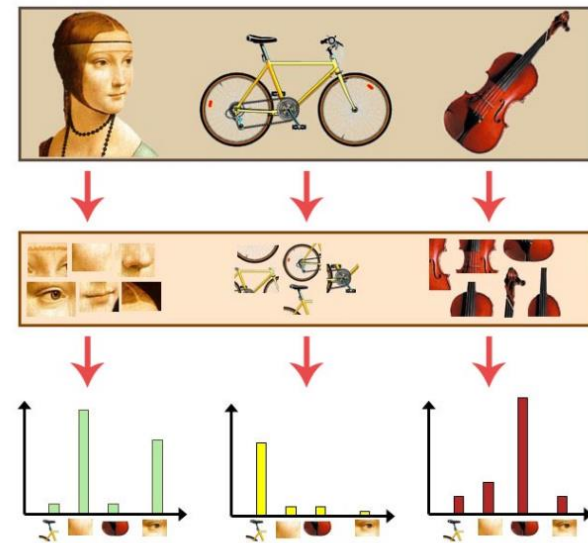
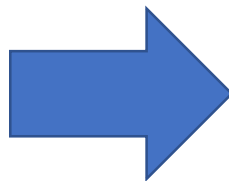
내용기반 이미지 검색 방법 (Content-based Image Retrieval)

1) 기존 → **BoVW 방법** 주로 사용



Keywords in documents

BoW기법 (Bag of Words) : 문서에서 각 단어의 사용 빈도수 (count)를 구하여 문서의 키워드가 무엇인지 파악하는 기법



Histogram of visual words

같은 개념으로, BoVW 기법은 단어 대신 이미지 특징점 (image features) 추출.
이미지 특징점 → 이미지에서 찾을 수 있는 unique pattern

1. 서론

즉, BoVW 방법 (Bag of Visual Words, 시각 단어집)

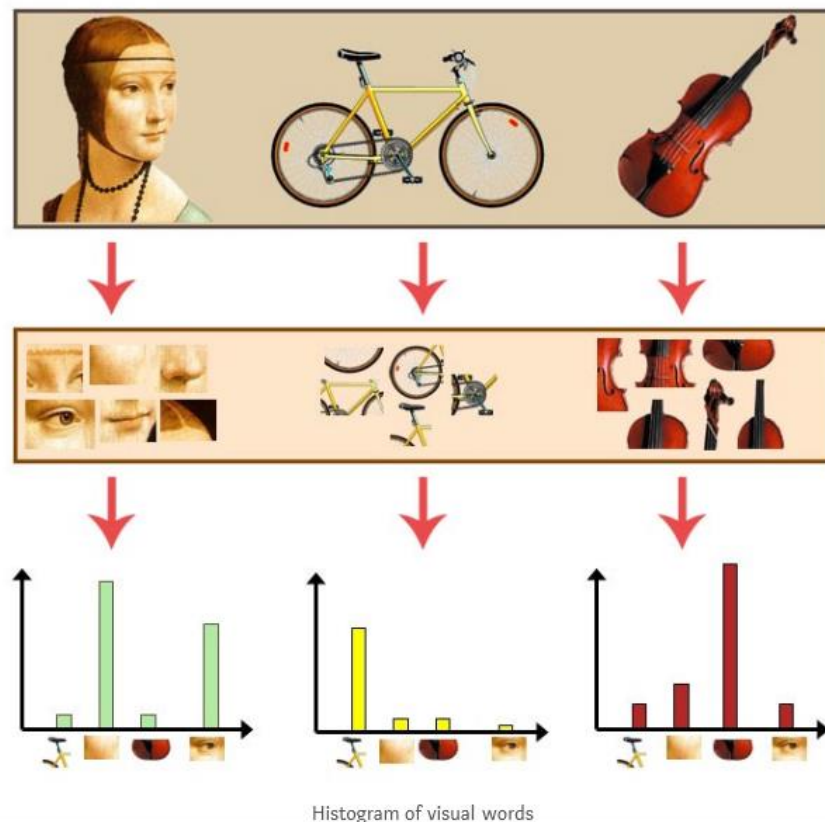
: SIFR(Scale Invariant Feature Transform)/ SURF(Speeded Up Robust Features)와 같은
특징점 추출 알고리즘 사용하여 이미지에서 변화에 강인한 **특징 추출** 후, 추출된 특징들로 군집화
과정(clustering)으로 단어집 구성. 구성된 시각 단어집으로 각 이미지의 내용정보 표현

특징 : keypoints , descriptors 로 구성

(1) Keypoints : "stand out" points in an image, so no matter the image is rotated, shrink, or expand, its keypoints will always be the same.


(2) descriptor : description of the keypoint

이 두 가지를 활용하여 어휘들(vocabularies)을 구성하고, 각 이미지를 이미지
속 특징들의 frequency histogram 로 나타냄. 이후의 frequency histogram을
이용하여 이미지 카테고리의 유사한 이미지를 찾거나 이미지 카테고리 예측



1. 서론

즉, BoVW 방법 (Bag of Visual Words, 시각 단어집)



: SIFT(Scale Invariant Feature Transform)/ SURF(Speeded Up Robust Features)와 같은
특징점 추출 알고리즘 사용하여 이미지에서 변화에 강인한 특징 추출 후, 추출된 특징들로 군집화
과정(clustering)으로 단어집 구성. 구성된 시각 단어집으로 각 이미지의 내용정보 표현

But, 이는 **물체 (object) 인식에는 성능 좋으나 배경 (scene) 이미지 분류에는 정확도가 떨어짐.**

Why? SIFT기반 BoVW는 **미분을 이용한 이미지 화소값의 변화 정도를 측정**, 그 변화가 큰 영역들을 뽑아내어
특징 벡터 구성
즉, 뚜렷한 경계를 지닌 객체 이미지에는 효과적이지만 배경 이미지에서는 X

2. 본론 (구현 및 실험)

내용기반 이미지 검색 방법 (Content-based Image Retrieval)

- 1) 기존 → BoVW 방법 주로 사용
- 2) 최근 → CNN 기반 딥러닝 기법

: 이미지에 컨볼루션 필터(convolution filter) 사용

- (a) 계산이 빠르고, 특정 객체 뿐만 아니라 배경을 포함한 이미지 전체를 고려할 수 있음
- (b) 별도의 특징 추출 알고리즘 없이, 특징 추출과 인식이 하나의 신경망을 통해 이루어지는 점이 다른 인식 알고리즘과 다른 특징

컨볼루션 신경망의 정의

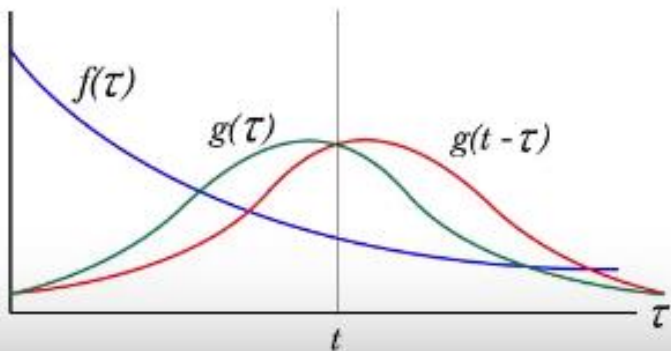
Definition of Convolution Neural Networks

1. 컨볼루션 신경망(CNN)의 정의

심층 신경망의 한 종류로, 하나 또는 여러 개의 컨볼루션 레이어와 통합 레이어(pooling layer), 완전 연결 레이어(fully connected layer) 등으로 구성된 이미지 분석에 특화된 신경망. 고양이의 시각과정에서 뇌의 시각 피질의 동작구조를 참고하여 1979년 쿠니히코 후쿠시마가 발표한 네오코그니트론(Neocognitron)모델이 발전한 형태.

2. 컨볼루션(convolution)

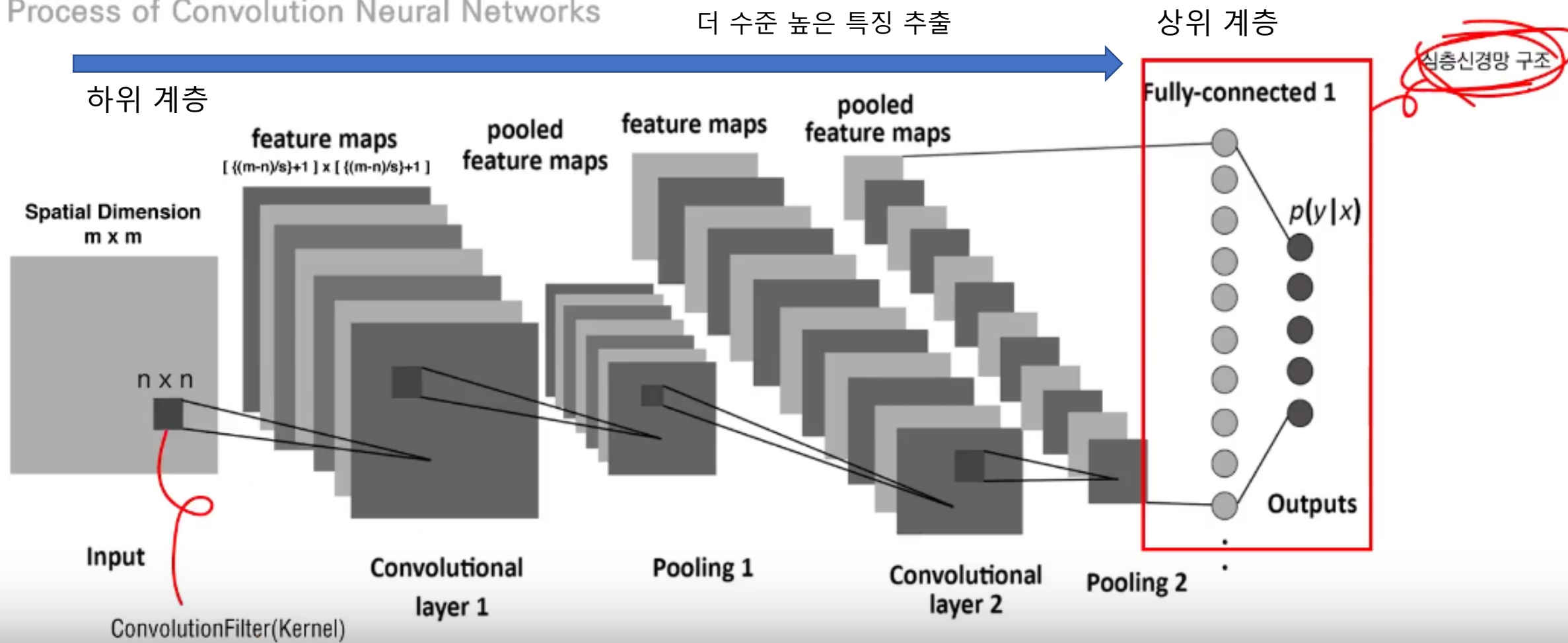
하나의 함수와 또 다른 함수를 반전 이동한 값을 곱한 다음, 구간에 대해 적분하여 새로운 함수를 구하는 연산자



$$f * g(t) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau = \int_{-\infty}^{\infty} g(\tau) f(t - \tau) d\tau$$

컨볼루션 신경망의 구조

Process of Convolution Neural Networks



컨볼루션 신경망의 구조

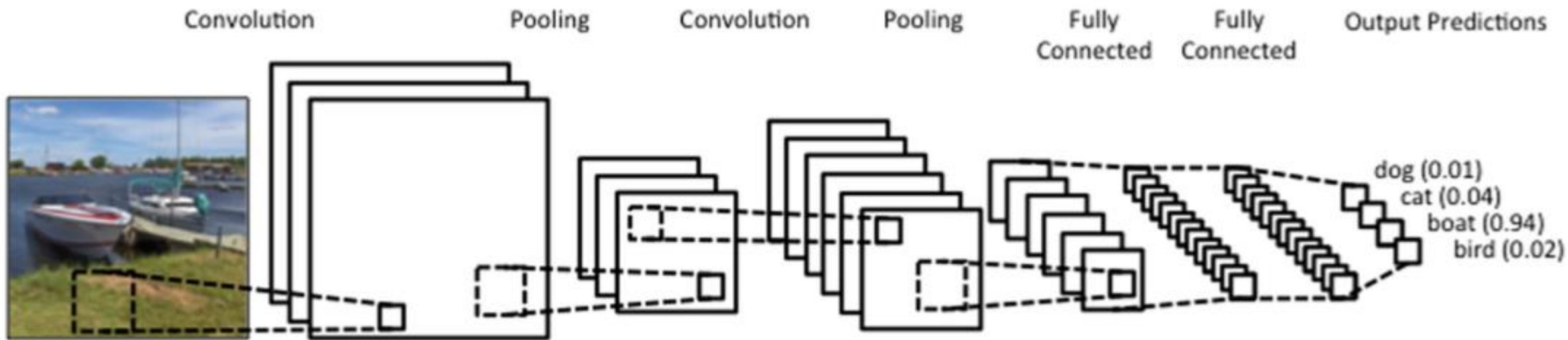


그림 1 Convolutional Neural Network 구조

Fig. 1 The architecture of CNN

스트라이드와 패딩

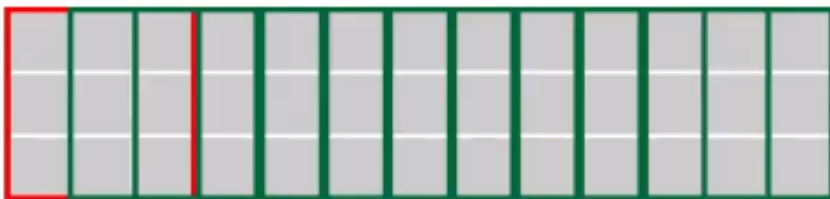
Stride and Padding

스트라이드[보폭](stride)

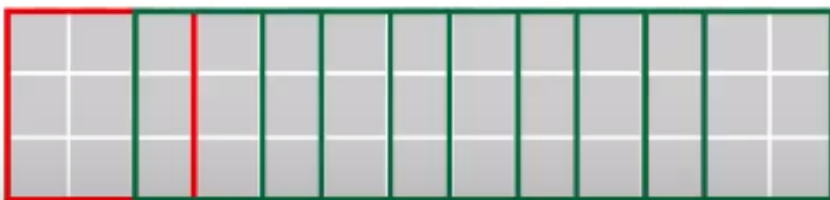
컨볼루션 필터가 합성곱 연산을 수행하기 위해 이동하는
일정한 이동량

On 3×3 convolution filter,

if stride=1



if stride=2

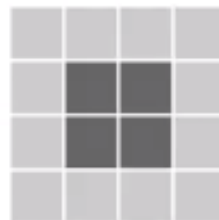


패딩(padding)

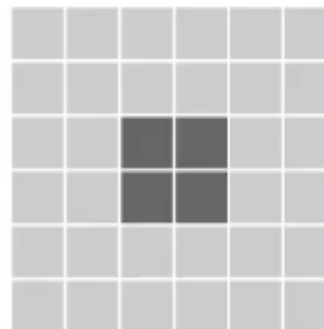
레이어 외부에 일정한 값의 레이어를 덧대는 것을 의미. 일반적으로
상하좌우 같은 크기와 값을 0으로 하는 패딩을 사용.

On 2×2 layer,

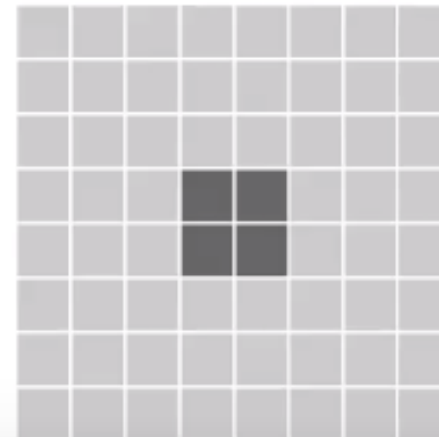
if padding=1



if padding=2

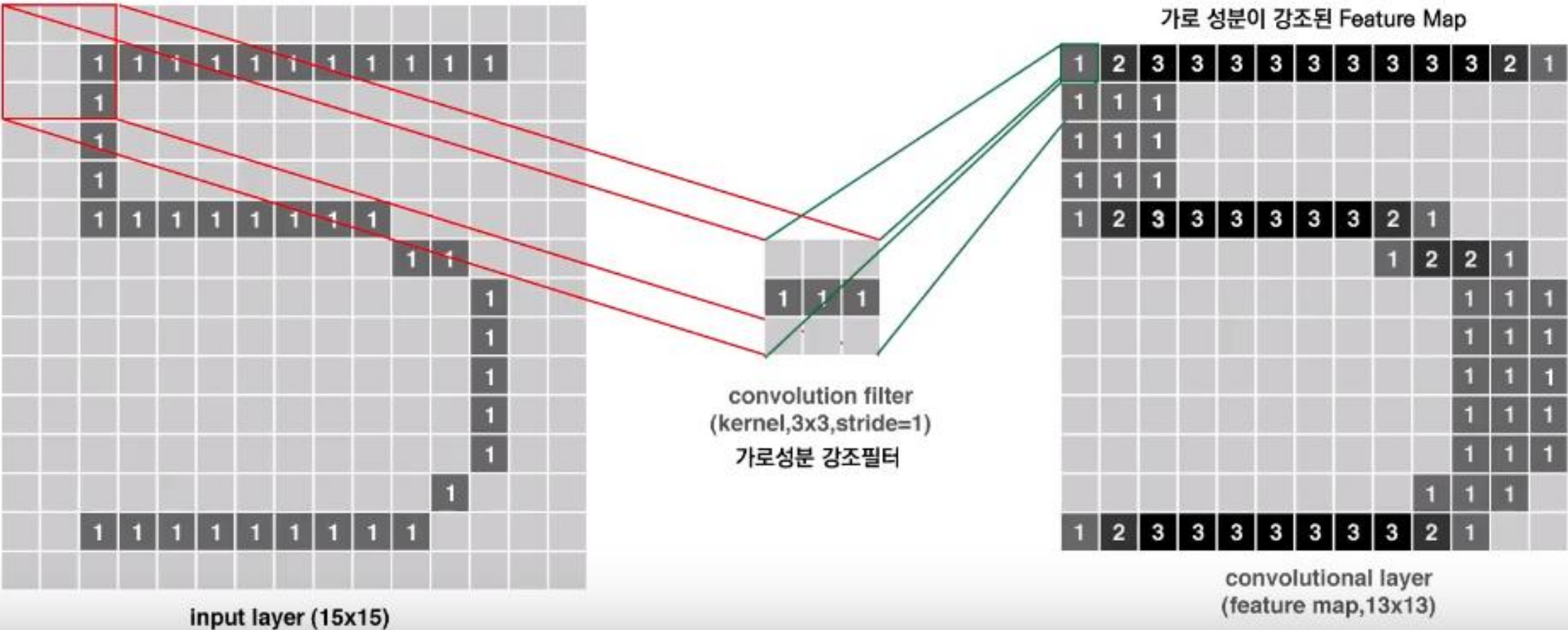


if padding=3



컨볼루션 필터(커널)의 적용

Application of Convolution Filter [Kernel]



논문에서의 구현

CNN 기반 훈련 → ImageNet 데이터셋 이용

테스트 → 인스타그램 15개의 카테고리 별 30개 이미지 = 총 450개의 인스타 이미지 테스트

$$\text{for } i, j = 0, 1, \dots, N-m \quad (4)$$

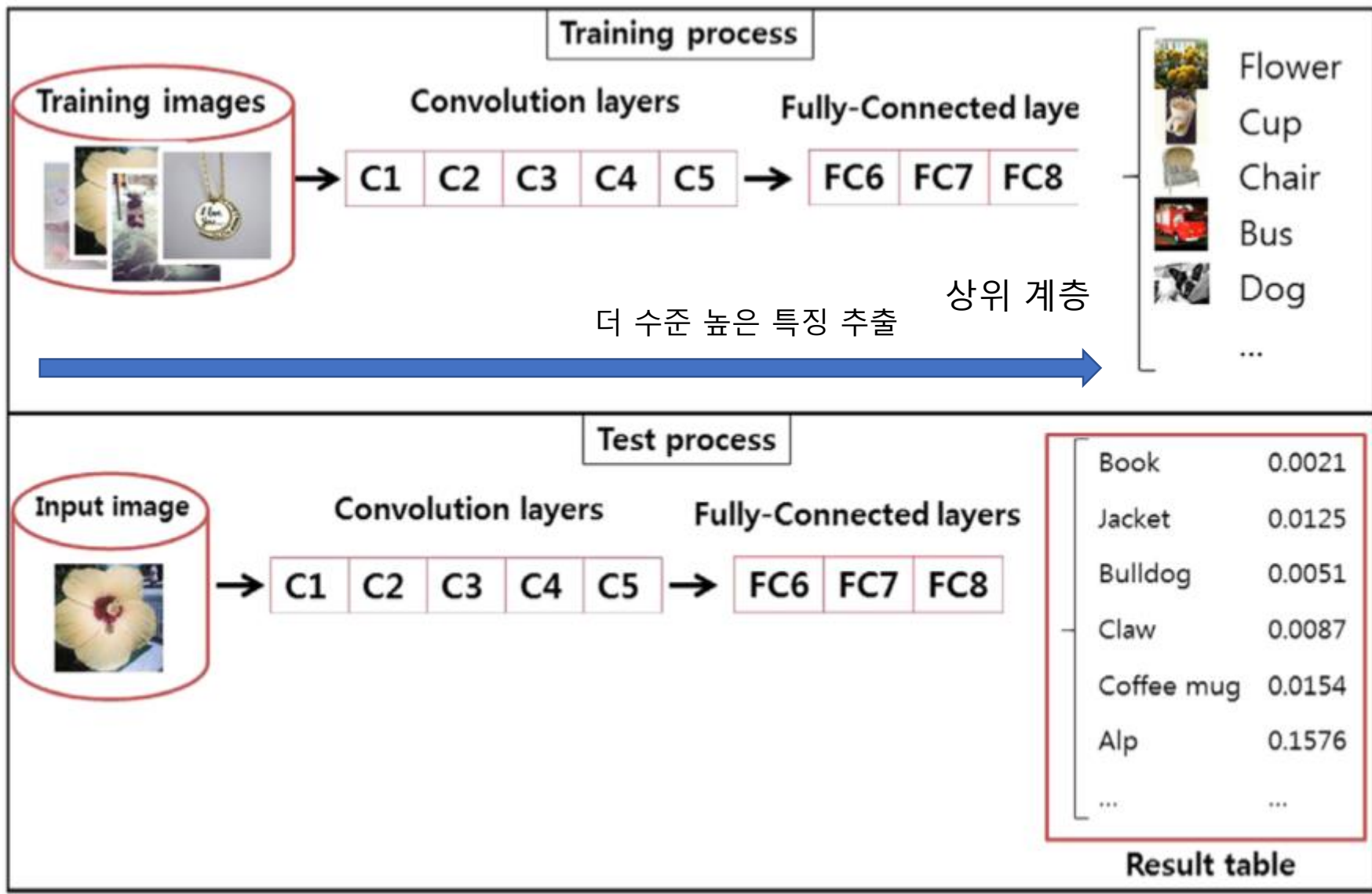
$$Z_{ij}^{(l)} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} y_{(i+a)(j+b)}^{(l-1)} \quad \text{Convolution filter로 가중치 계산}$$

$$y_{ij}^{(l)} = f(Z_{ij}^{(l)}), \text{ } f \text{ is nonlinear function} \quad (5) \quad \text{활성화 함수로 계산}$$

$$f(x) = \max(0, x), \quad f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (6) \quad (\text{ReLU 함수 이용})$$

2. 본론

하위 계층




Softmax 함수로
최종적인 분류
수행

그림 3 전체적인 CNN 구조
Fig. 3 The architecture of CNN

최종 구현 결과 (1)

표 1 인스타그램 태그와 CNN을 통해 태깅한 결과 비교

Table 1 A comparison of tagging results using CNN and instagram tags

| 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 |

→ 인스타그램 태그보다 CNN으로 자동 태깅 된 결과가 더 섬세하고 정확

최종 구현 결과 (2)

표 2 CNN을 사용해서 태깅한 결과의 검색 정확도 향상

Table 2 Improvement of retrieval correctness from tagging results using CNN

| Categories | SIFT | | instagram | | CNN | |
|------------|--------------|-----------|--------------|-----------|--------------|-----------|
| | correct tags | Rates (%) | correct tags | Rates (%) | correct tags | Rates (%) |
| bike | 3 | 10.00 | 20 | 66.67 | 23 | 76.67 |
| bird | 20 | 66.67 | 15 | 50.00 | 21 | 70.00 |
| bottle | 15 | 50.00 | 23 | 76.67 | 21 | 70.00 |
| building | 3 | 10.00 | 23 | 76.67 | 15 | 50.00 |
| bus | 6 | 20.00 | 8 | 26.67 | 21 | 70.00 |
| car | 19 | 63.33 | 23 | 76.67 | 28 | 93.33 |
| chair | 5 | 16.67 | 19 | 63.33 | 18 | 60.00 |
| cup | 18 | 60.00 | 19 | 63.33 | 27 | 90.00 |
| flower | 19 | 63.33 | 18 | 60.00 | 18 | 60.00 |
| food | 6 | 20.00 | 12 | 40.00 | 22 | 73.33 |
| house | 10 | 33.33 | 7 | 23.33 | 17 | 56.67 |
| mountain | 3 | 10.00 | 17 | 56.67 | 21 | 70.00 |
| pet | 15 | 50.00 | 24 | 80.00 | 22 | 73.33 |
| roadsign | 22 | 73.33 | 27 | 90.00 | 22 | 73.33 |
| sea | 1 | 3.33 | 19 | 63.33 | 21 | 70.00 |
| Total | 165 | 36.67 | 274 | 60.89 | 317 | 70.44 |

→ SIFT < 인스타그램 태그 < CNN으로 자동 태깅

감사합니다 😊