

Grad-CAM:

Visual Explanations from Deep Networks via Gradient-based
Localization

Tae-hwi Kim

ICCV, 2017

Preliminaries

- 대부분의 딥러닝 모델은 예측 결과에 대한 원인을 쉽게 파악할 수 없음
- 하지만 예측에 대한 신뢰성과 추가적인 분석을 위해 그 원인을 파악하는 작업은 매우 중요하다.

> CAM(Class Activation Mapping)

: **Global Average Pooling (GAP) Layer**를 이용해 CNN의 예측 결과에 대해 그 원인을 파악할 수 있는 방법을 제시



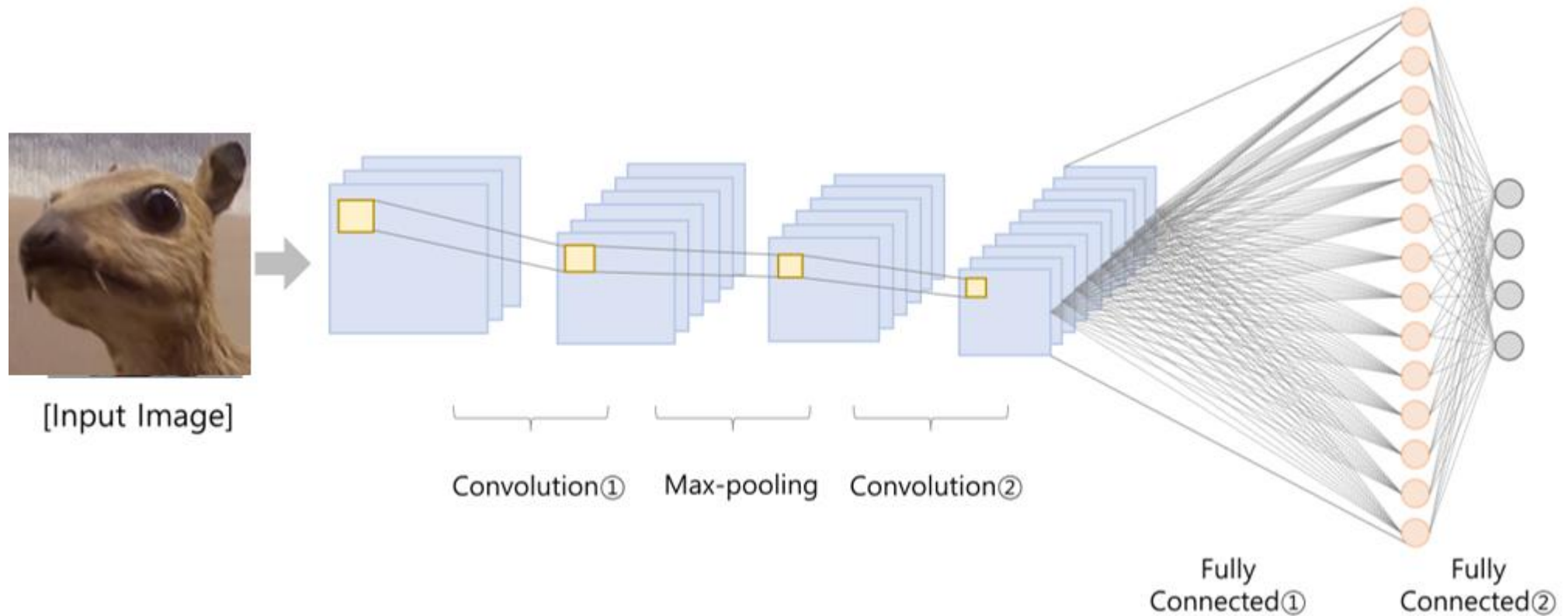
<고관절 골절 탐지>
→ 병 원인 진단

Preliminaries

- *General CNN*

각 Feature map을 Flatten하는 구조

* Images from..
<https://tyami.github.io/deep%20learning/CNN-visualization-Grad-CAM/>

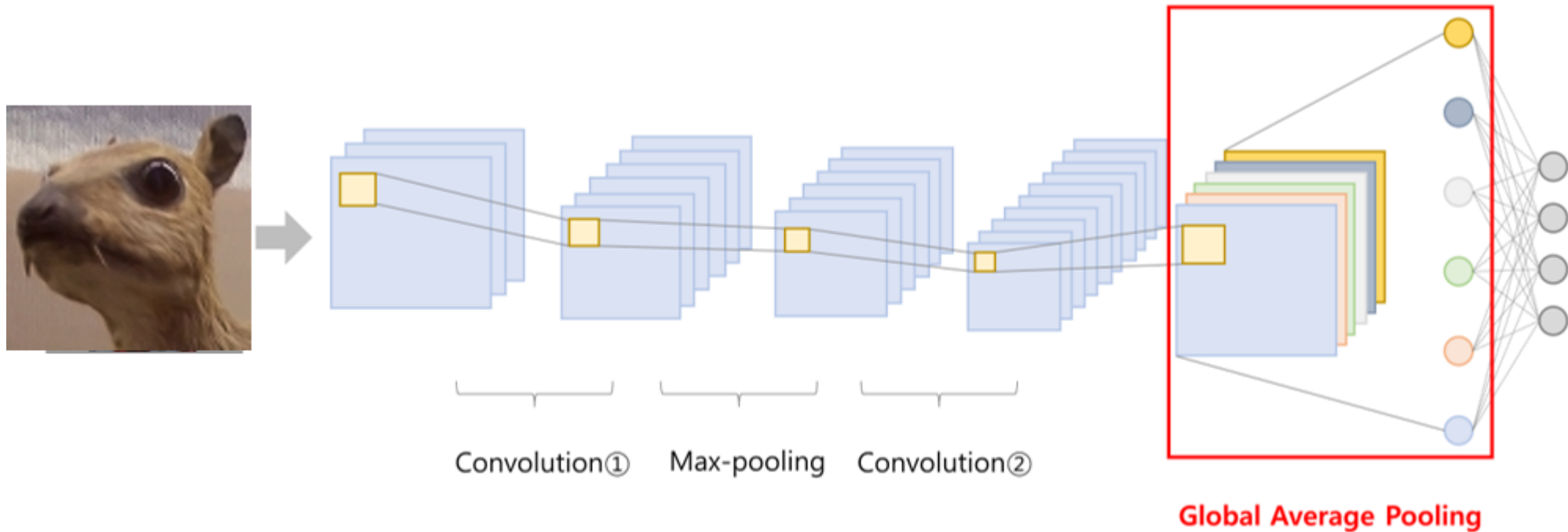


Preliminaries

- *CAM*

: 기존 CNN 마지막 Layer와 FC Layer 사이에
GAP Layer를 추가!

* Images from..
<https://tyami.github.io/deep%20learning/CNN-visualization-Grad-CAM/>

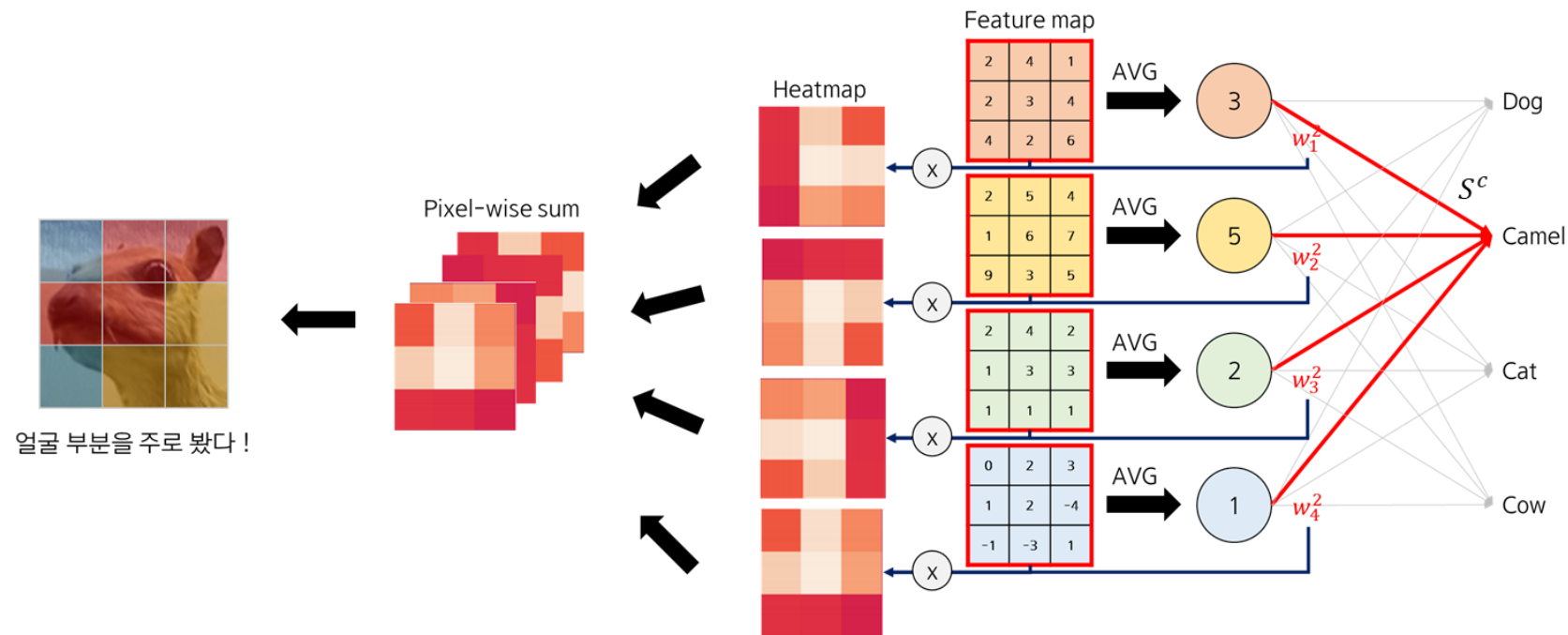


Preliminaries

- CAM*

* Images from..

<https://tyami.github.io/deep%20learning/CNN-visualization-Grad-CAM/>



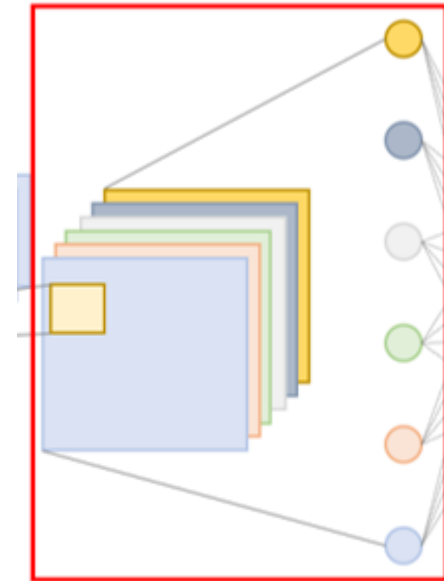
$$L_{CAM}^c(i, j) = \sum_k w_k^c f_k(i, j)$$

Preliminaries

- *CAM - Limitations*

1. *Global Average Pooling Layer가 반드시 필요하다.*
2. *Conv layer의 마지막 Layer에서만 사용할 수 있다.*

› 좀 더 괜찮은 방법이 있을까?

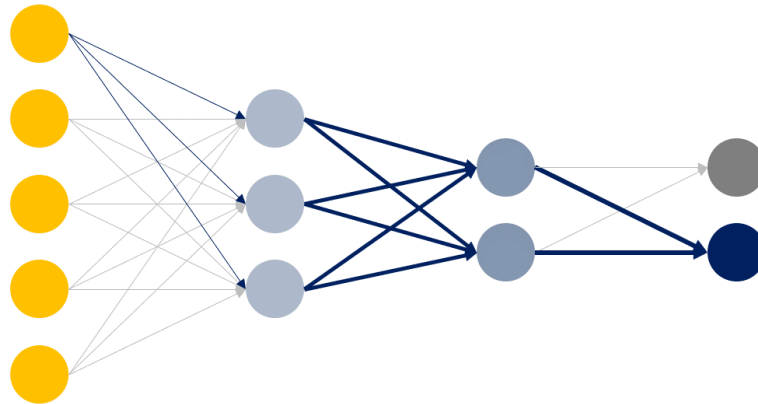


Global Average Pooling

Grad-CAM

- *Introduction*

CAM의 한계점을 보완하기 위해 GAP Layer를 사용하지 않는 방법을 제안,
“Gradient는 Input x 가 어떤 Class C 에 대해 끼치는 영향력”이라는 관점에서 시작.



Gradient
[Total amount of effect of input K on output class C]

Grad-CAM

- Methods*

기존의 CNN 구조를 그대로 사용하며, GAP을 통해 얻는 Feature map의 Weight는 Gradient를 통해 얻는다.

+) Pixel-wise sum 결과에 대한 ReLU 함수 적용.

<C class에 대한 CAM Score>

$$S^c = \sum_k W_k^c \frac{1}{Z} \sum_i \sum_j A_{i,j}^k$$



< C class에 대한
Grad_CAM Score >

$$L_{Grad-CAM}^c = ReLU \sum_k a_k^c A^k$$

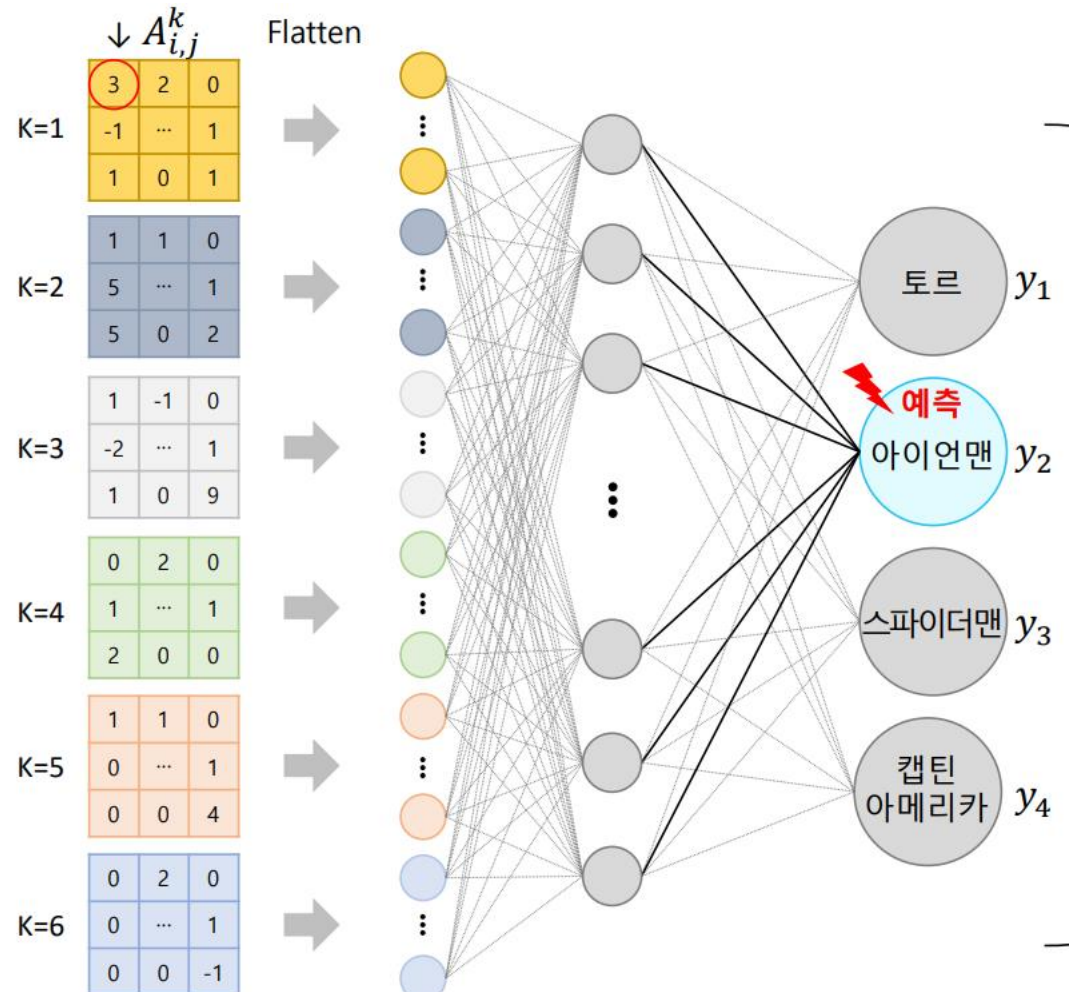
$$a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_{i,j}^k}$$

Grad-CAM

- Methods*

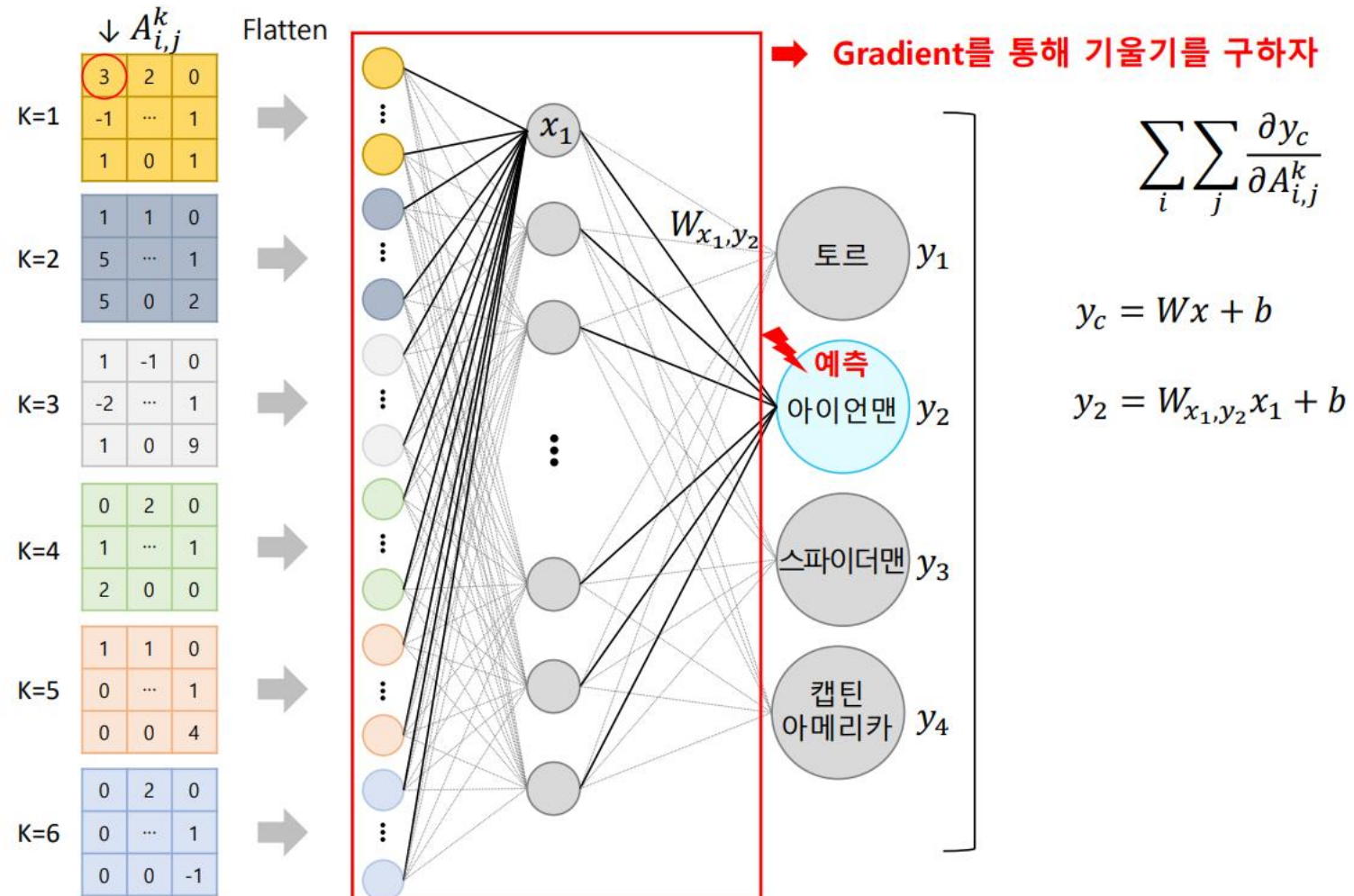
기존 CNN 구조를 그대로 사용,

각 Feature map을 Flatten 해준다.



Grad-CAM

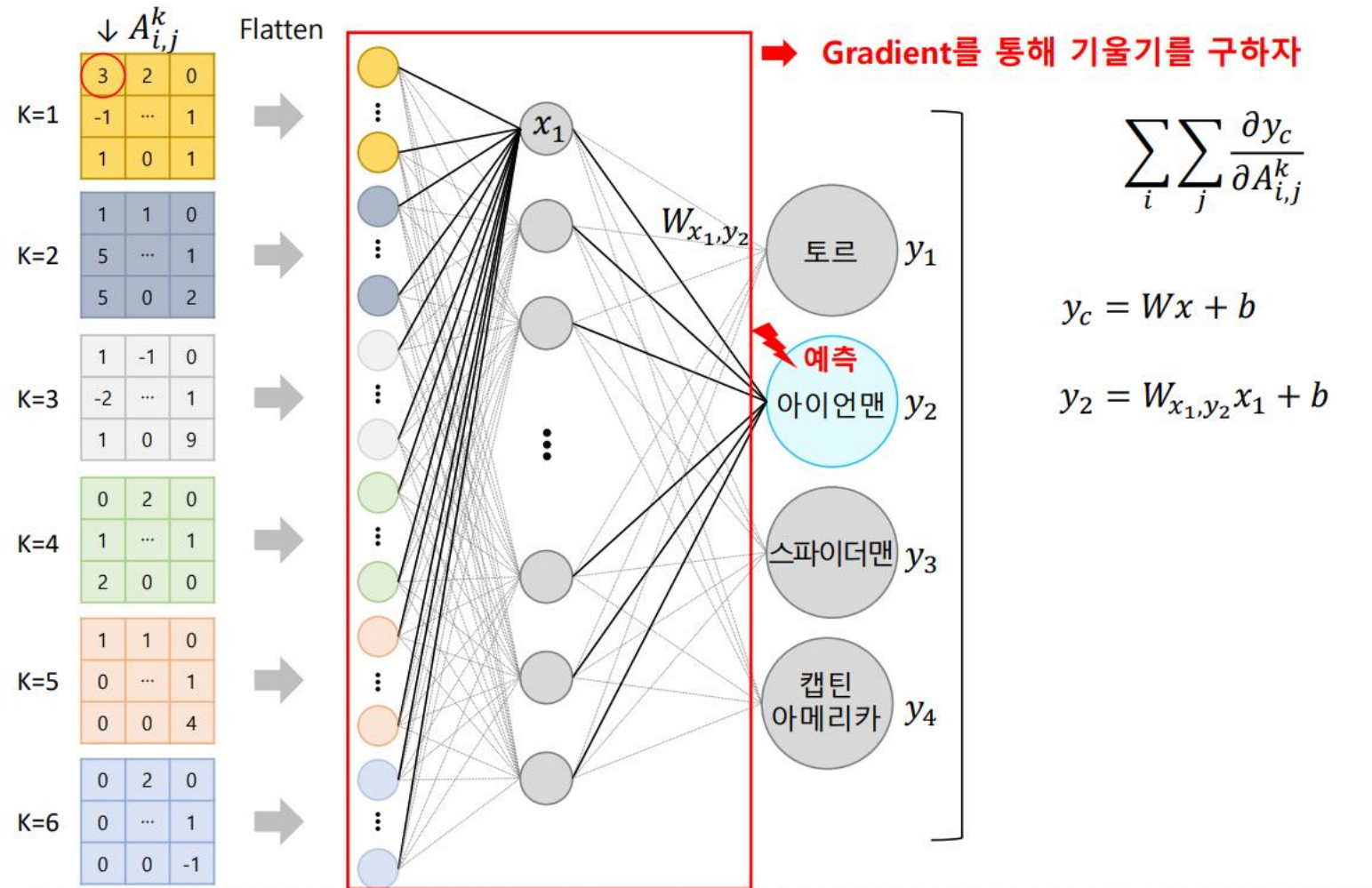
• Methods



Grad-CAM

• Methods

> Class 예측 결과에 대해,
각 Feature map의
(i,j)에 대한 **Gradient**를
Weight로 사용하자.

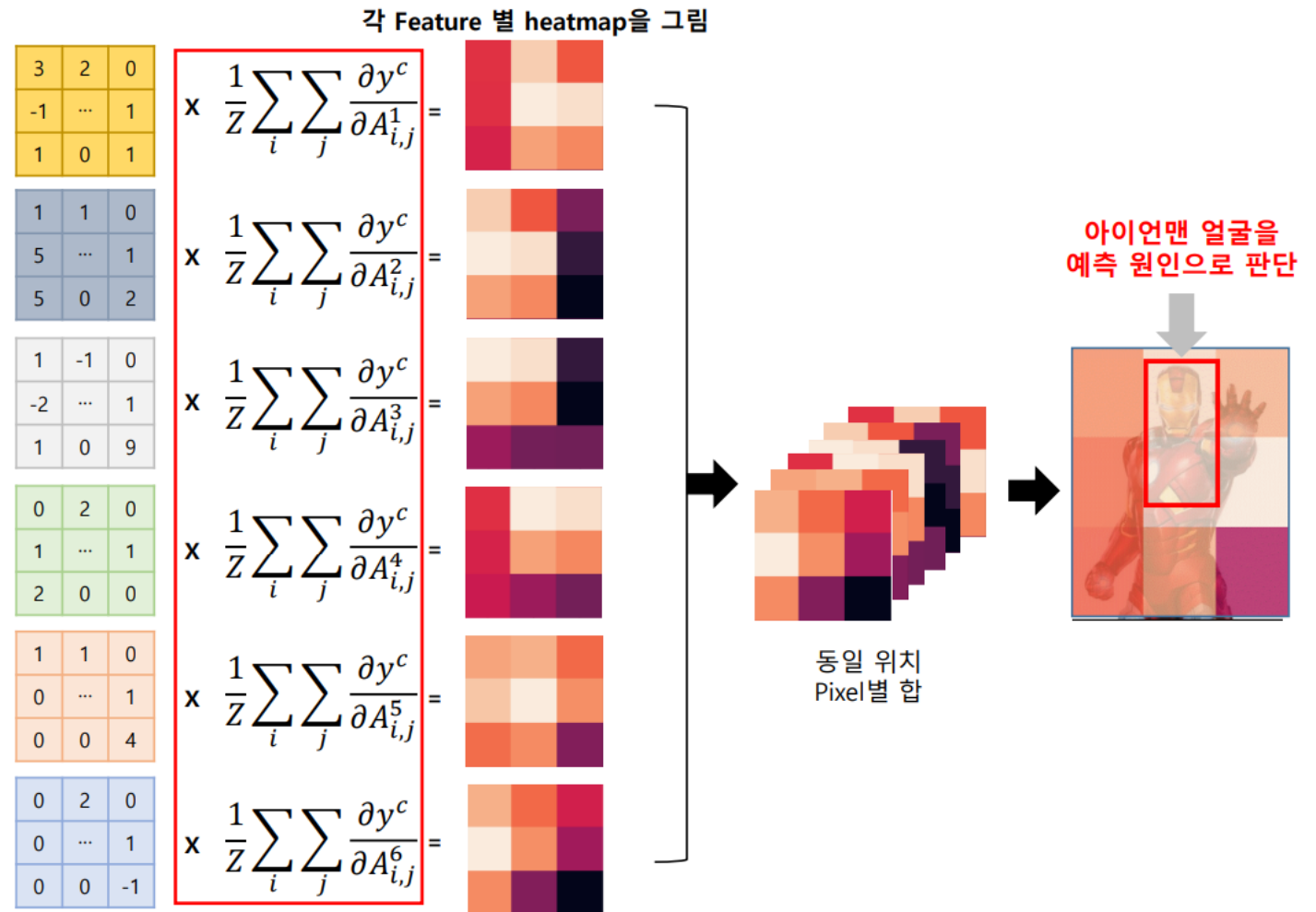


Grad-CAM

• Methods

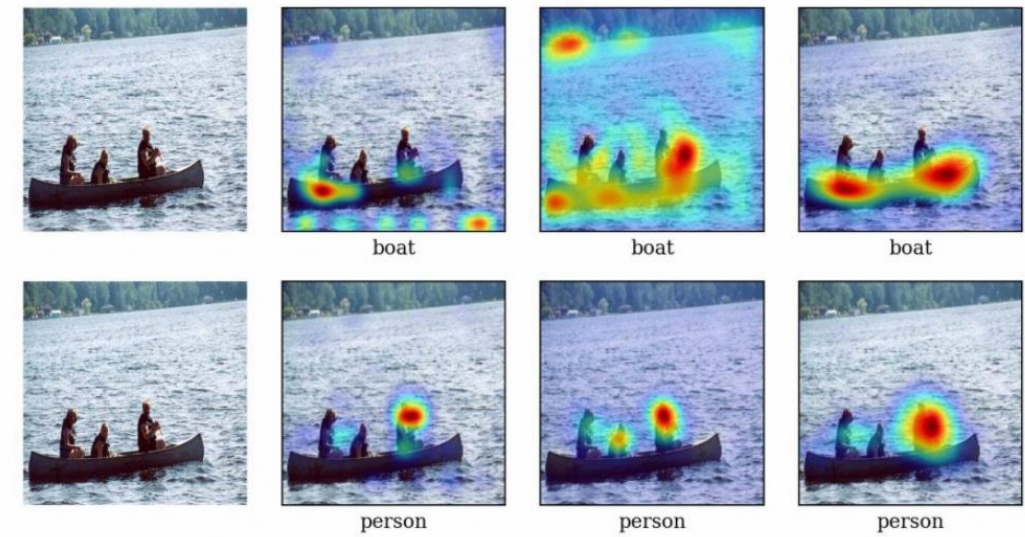
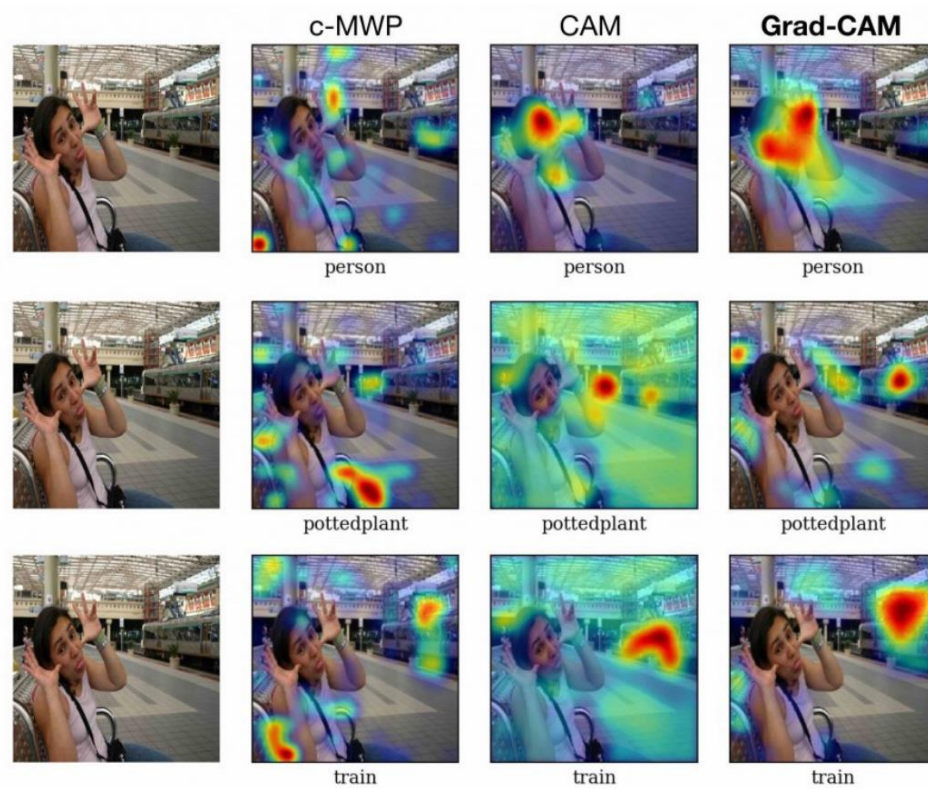
Feature map에 GAP을 통한 Weight가 아닌 Gradient로 Heatmap 생성.

> 여기서 Z는 각 Feature map의 합으로, 해당 픽셀(i,j)이 전체에 대해 비율 상 얼마나 영향력을 가지는 지 나타나게 해줌.



Grad-CAM

- *Comparison with CAM*



Grad-CAM

- Experimental Results*

		Classification		Localization	
		Top-1	Top-5	Top-1	Top-5
VGG-16	Backprop [51]	30.38	10.89	61.12	51.46
	c-MWP [58]	30.38	10.89	70.92	63.04
	Grad-CAM (ours)	30.38	10.89	56.51	46.41
	CAM [59]	33.40	12.20	57.20	45.14
AlexNet	c-MWP [58]	44.2	20.8	92.6	89.2
	Grad-CAM (ours)	44.2	20.8	68.3	56.6
GoogLeNet	Grad-CAM (ours)	31.9	11.3	60.09	49.34
	CAM [59]	31.9	11.3	60.09	49.34

Table 1: Classification and localization error % on ILSVRC-15 val (lower is better) for VGG-16, AlexNet and GoogLeNet. We see that Grad-CAM achieves superior localization errors without compromising on classification performance.

Grad-CAM

- *Summary*

1. 모델 구조의 변형 없이 예측 결과에 대한 분석 및 원인 파악 가능.
2. Gradient(by back-propagation)를 사용하기 때문에 어느 Layer에서든 heatmap을 그려 분석할 수 있음.
3. ReLU를 사용하면서 보다 많은 Class에 대해 명확한 Localization이 이루어지는 모습을 확인할 수 있었음.