History of
Class Activation Map (CAM)

발표자 : 백인성

2022.02.25.

Data Mining
Quality Analytics

발표자 소개

Insung Baek

❖ 백인성 (Insung Baek)

- Korea University
- Data Mining & Quality Analytics Lab
- Ph.D. Candidate (2018. 9 ~ Present)

❖ Research Interest

- Explainable Artificial Intelligence algorithms (Explainable A.I.)
- Game Artificial Intelligence (Game A.I.)
- Application of deep learning and machine learning algorithms.

❖ Contact

- E-mail: insung_baek01@korea.ac.kr

목차

1. Introduction

2. Class Activation Map(CAM)

3. Gradient based CAM

4. Score-CAM

5. LFI-CAM

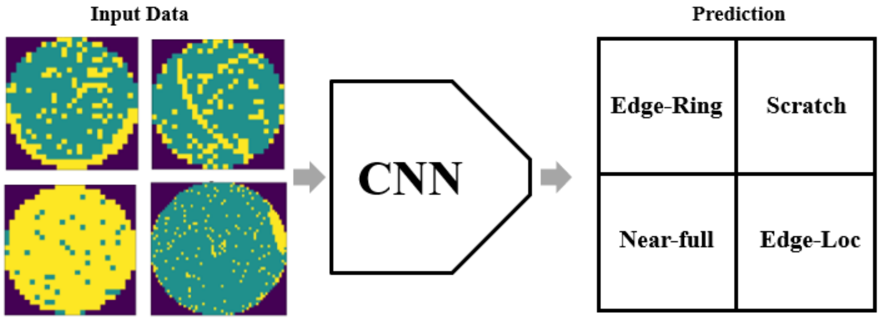
6. Conclusions

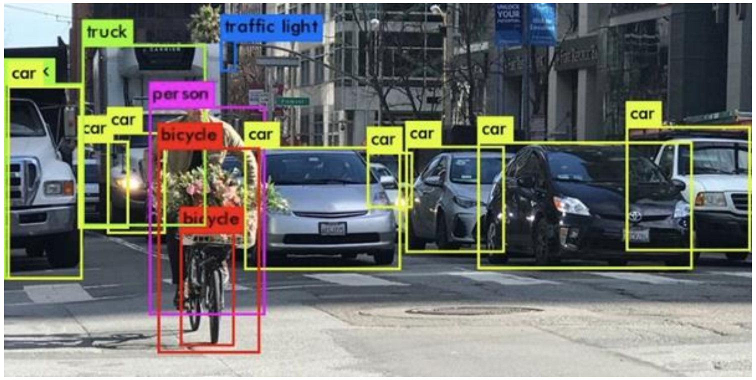
1. Introduction

Introduction

이미지 분야에서 인공지능의 발전

- ❖ 이미지 분야의 다양한 문제를 푸는 인공지능이 빠르게 발전하고 있다.
- ❖ 제조업에서 이미지를 통해 불량의 유형을 자동 탐지해 비용을 절감하는 노력이 진행됨
- ❖ 자동 객체 탐지를 통해 자율 주행 차량의 발전도 진행되고 있음





<제조업에서 불량 유형 자동 탐지>

<자율 주행 차량을 위한 객체 탐지>

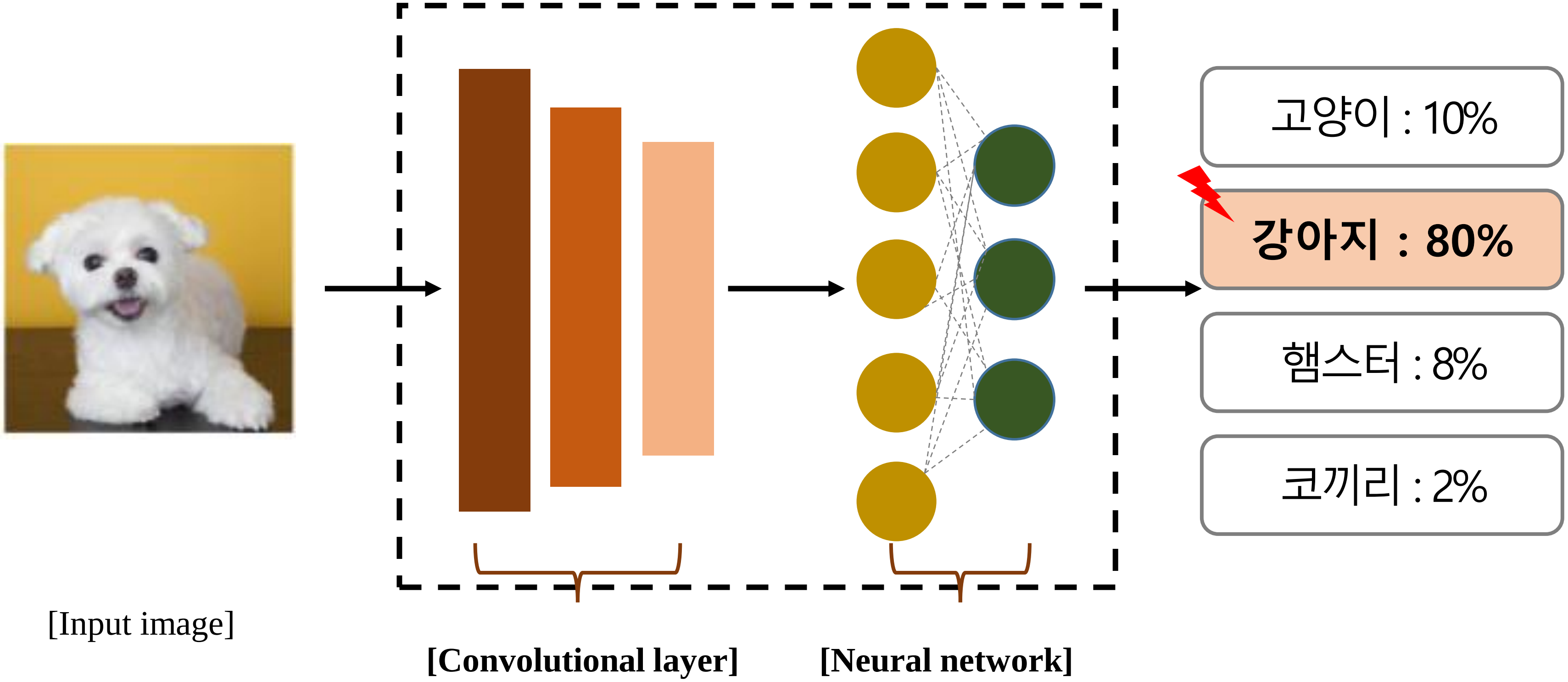
Reference: https://www.datamaker.io/posts/17/

Introduction

Convolutional Neural Network (CNN)

❖ Convolutional Neural Network 기본 구조

- Neural network 모델에 convolution layer를 사용한 방법론
- Classification (분류), object detection (객체 탐지) 등 visual task에서 좋은 성능을 보임

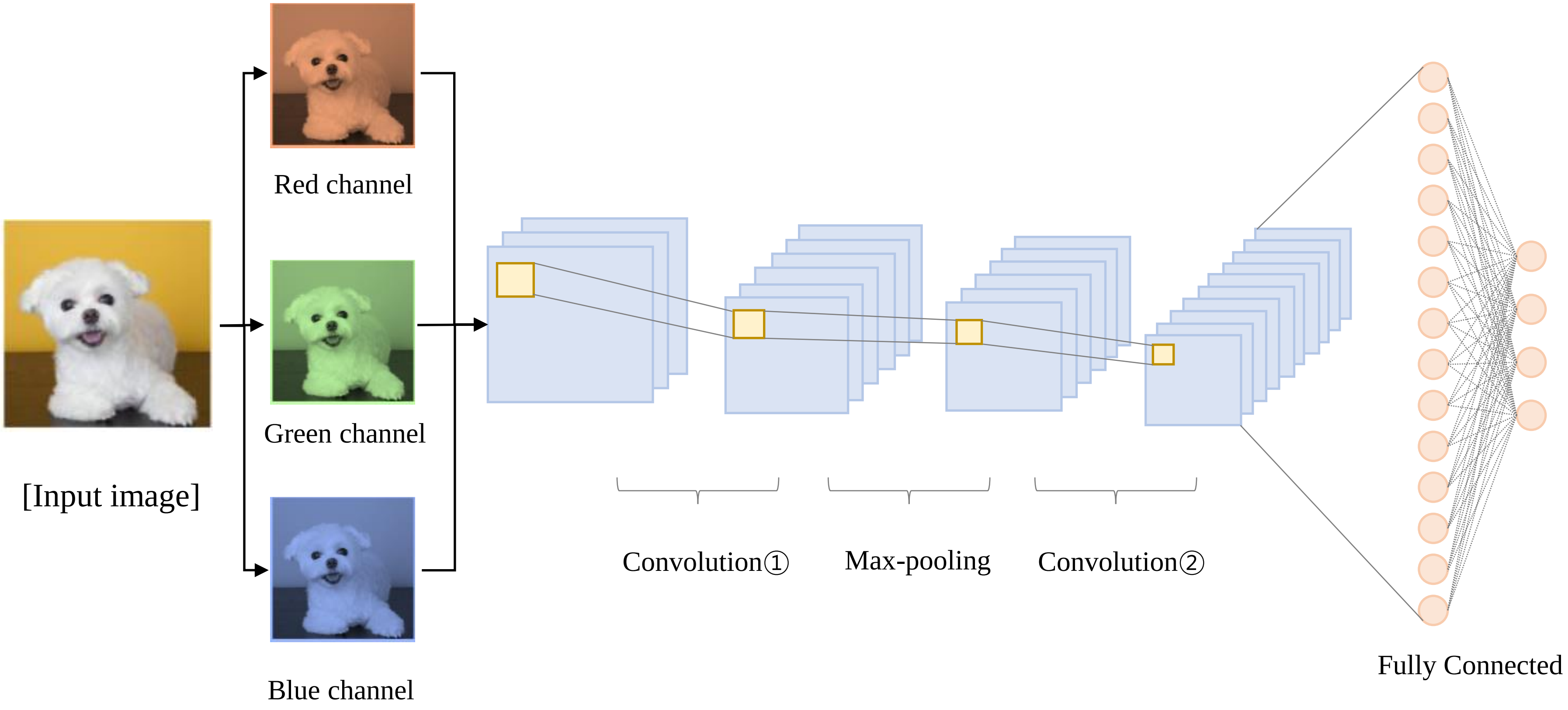


Introduction

Convolutional Neural Network (CNN)

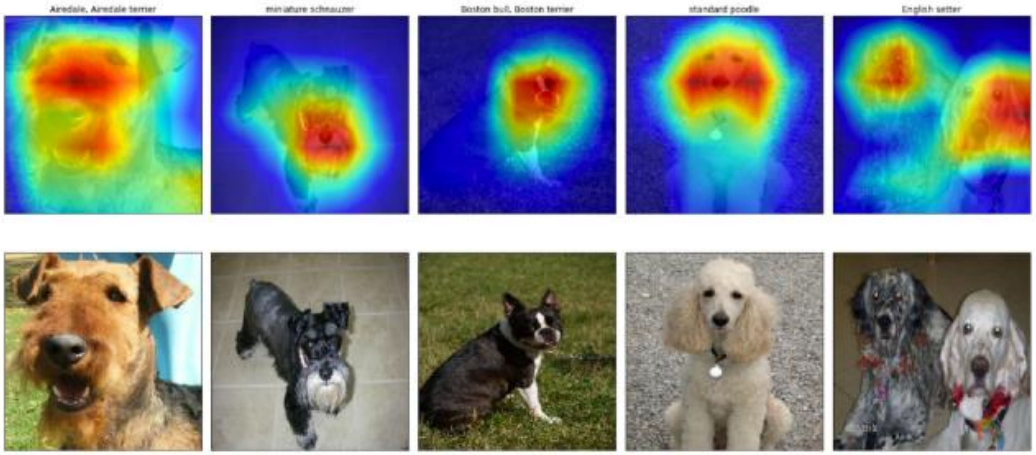
❖ Convolutional Neural Network 세부 구조

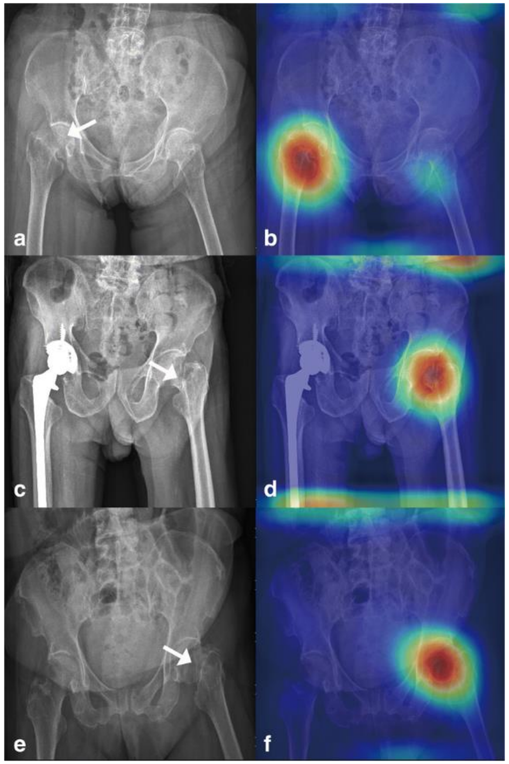
- 입력 이미지에 대해 convolution & pooling layer를 활용하여 특징을 추출함
- 분류 모델에서는 입력 이미지가 어떤 class에 속하는지 예측함



Introduction

❖ 예측 모델에서 실제 CAM(Class Activation Map)을 활용한 사례





<분류 모델 원인 해석>
→ 예측 모델 결과에 대한 신뢰성

<고관절 골절 탐지>
→ 병 원인 진단

Reference: https://alexisbcook.github.io/2017/global-average-pooling-layers-for-object-localization/
Cheng, Chi-Tung, et al. "Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs." European radiology (2019): 1-9.

Introduction

❖ 예측 모델의 해석이 필요한 이유는 아래의 3가지 이유로 요약

1. 예측 모델을 사용하는 현업자들에게 이해하기 쉽도록

2. 구축한 예측 모델 결과가 타당함을 직관적으로 보이기 위해

3. 예측 결과에 대한 원인을 분석하고 향후 대처할 수 있게

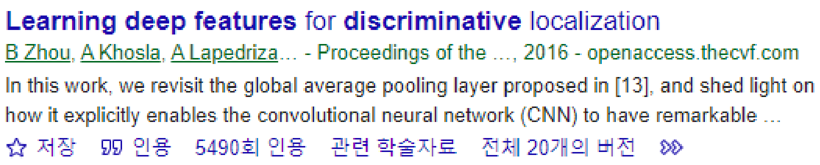
2. Class Activation Map(CAM)

Class Activation Map (CAM)

CAM 논문

❖ Class Activation Map (CAM) (2016)

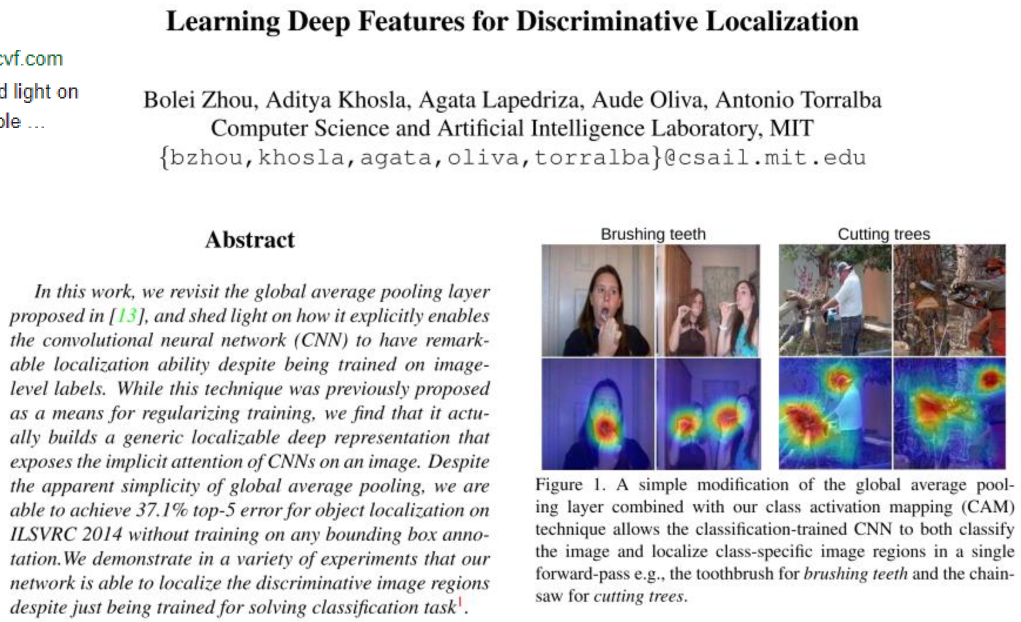
- 딥러닝 프레임 워크에서 예측 원인을 파악하기 위해 등장한 방법론
- 2016년도 CVPR (Computer Vision and Pattern Recognition)에서 발표됨
- 22년 2월 24일 기준 5,409회 인용



Learning deep features for discriminative localization

B Zhou, A Khosla, A Lapedriza… - Proceedings of the …, 2016 - openaccess.thecvf.com

In this work, we revisit the global average pooling layer proposed in [13], and shed light on how it explicitly enables the convolutional neural network (CNN) to have remarkable …



Learning Deep Features for Discriminative Localization

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba
Computer Science and Artificial Intelligence Laboratory, MIT
{bzhou,khosla,agata,oliva,torralba}@csail.mit.edu

Abstract

In this work, we revisit the global average pooling layer proposed in [13], and shed light on how it explicitly enables the convolutional neural network (CNN) to have remarkable localization ability despite being trained on image-level labels. While this technique was previously proposed as a means for regularizing training, we find that it actually builds a generic localizable deep representation that exposes the implicit attention of CNNs on an image. Despite the apparent simplicity of global average pooling, we are able to achieve 37.1% top-5 error for object localization on ILSVRC 2014 without training on any bounding box annotation. We demonstrate in a variety of experiments that our network is able to localize the discriminative image regions despite just being trained for solving classification task[1].

Figure 1. A simple modification of the global average pooling layer combined with our class activation mapping (CAM) technique allows the classification-trained CNN to both classify the image and localize class-specific image regions in a single forward-pass e.g., the toothbrush for *brushing teeth* and the chainsaw for *cutting trees*.

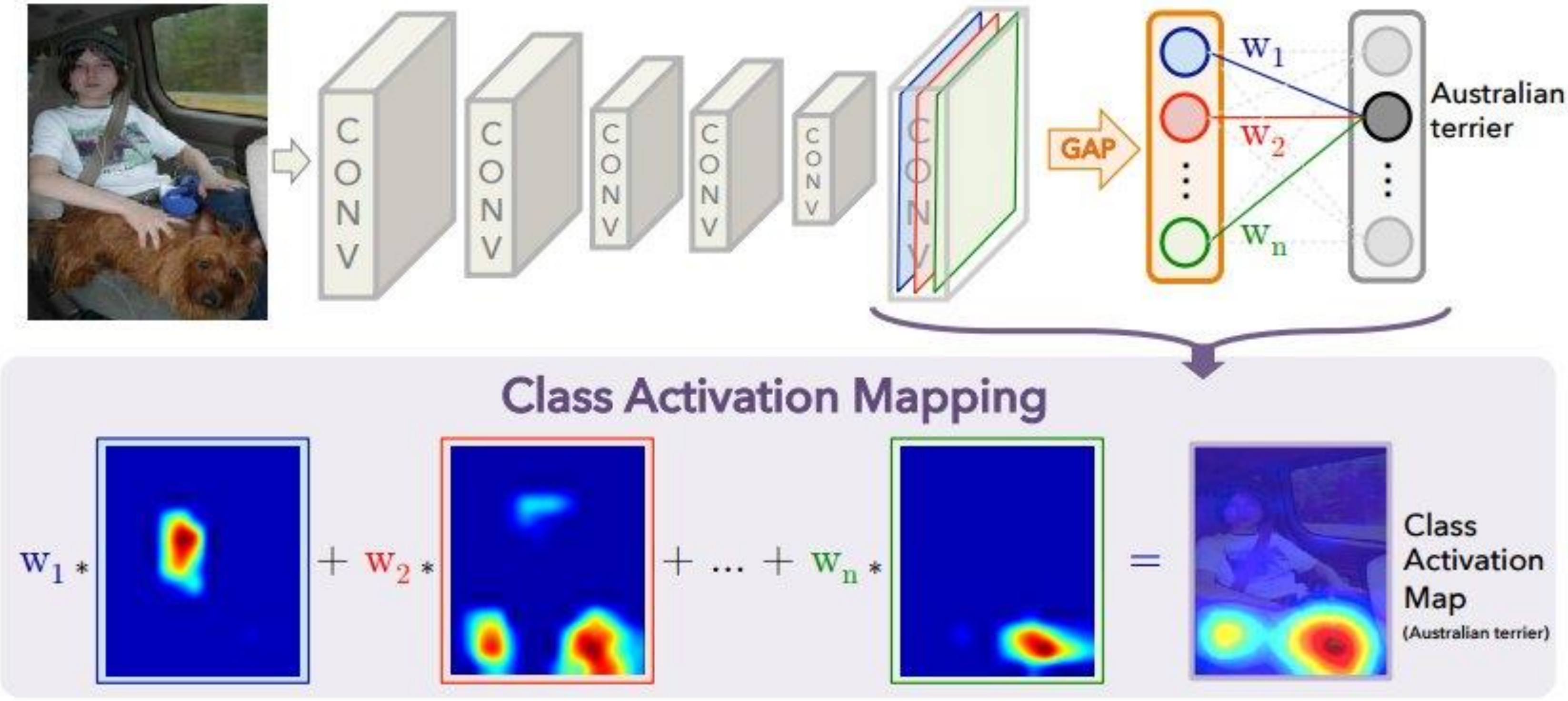
Reference: Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization.
In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

Class Activation Map (CAM)

CAM 알고리즘 요약

❖ Class Activation Map 구조

- CNN 모델로 예측 시, 어떤 부분이 class 예측에 큰 영향을 주었는지 확인 가능
- 마지막 Convolutional layer 이후 Global Average Pooling (GAP) 사용



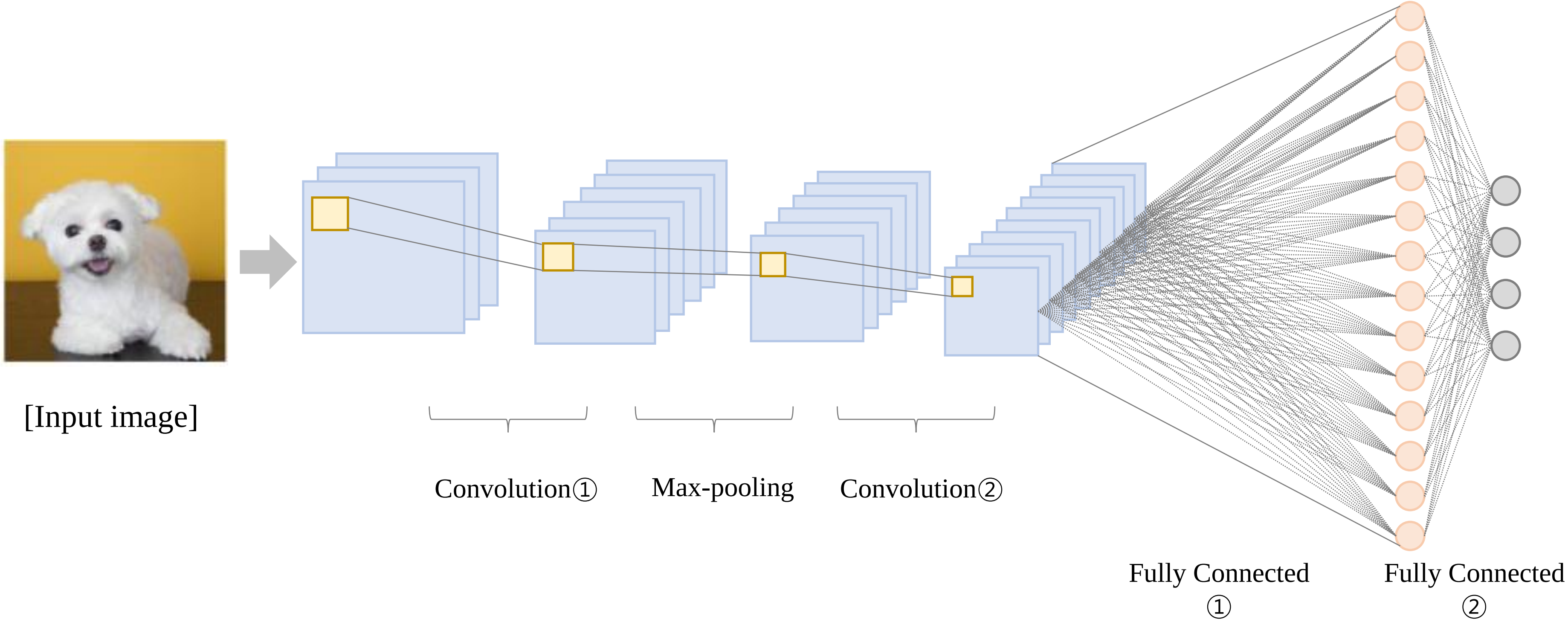
출처: Zhou, Bolei, et al. "Learning deep features for discriminative localization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Class Activation Map (CAM)

일반적인 CNN 구조

❖ Convolutional Neural Network(CNN) 구조

- 이미지를 입력 변수로 활용해 이미지의 class를 맞추는 분류 모델 구조
- Convolution layer와 pooling layer를 활용해서 이미지 내 정보를 요약
- 최종 분류 예측 전에 Fully connected layer 활용해 정답 class를 예측함

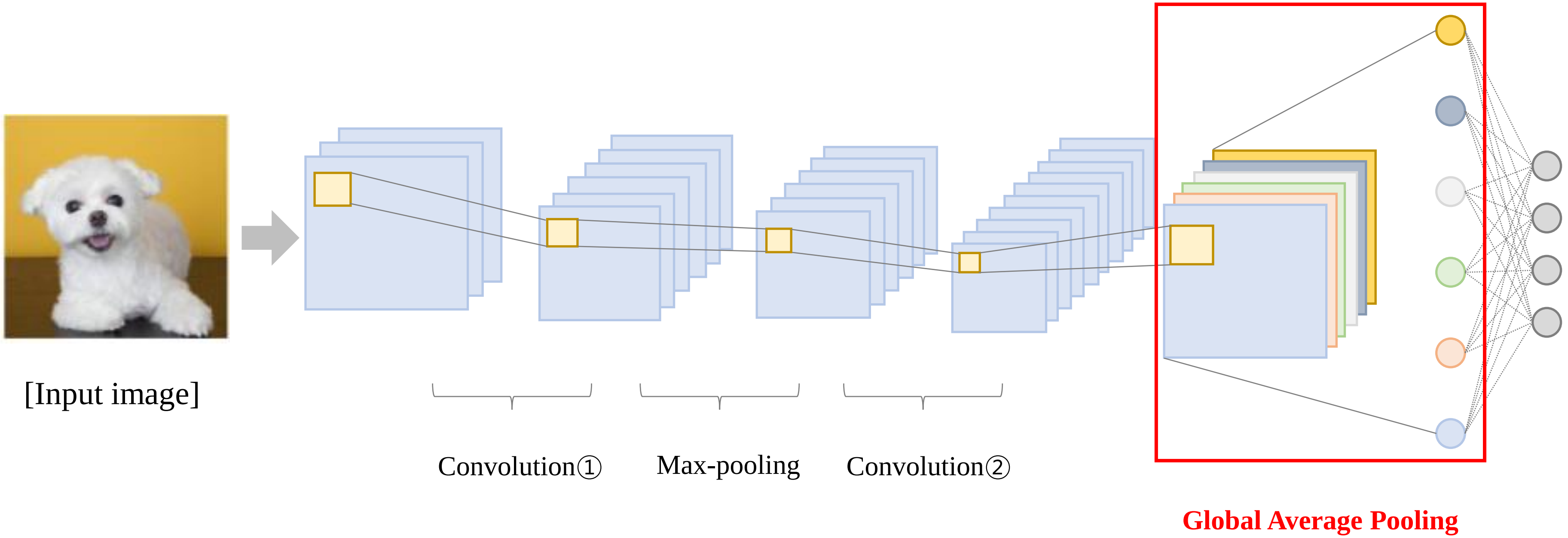


Class Activation Map (CAM)

Global Average Pooling (GAP) 구조

❖ CNN + Class Activation Map(CAM) 구조

- Convolution layer와 pooling layer를 활용해서 이미지 내 정보를 요약
- 마지막 Convolutional layer 뒤에 Global Average Pooling (GAP)구조를 사용



Class Activation Map (CAM)

Global Average Pooling → 각 Feature별 평균 값을 구함

[Input image]

[Last feature map]

3	2	0
-1	…	1
1	0	1

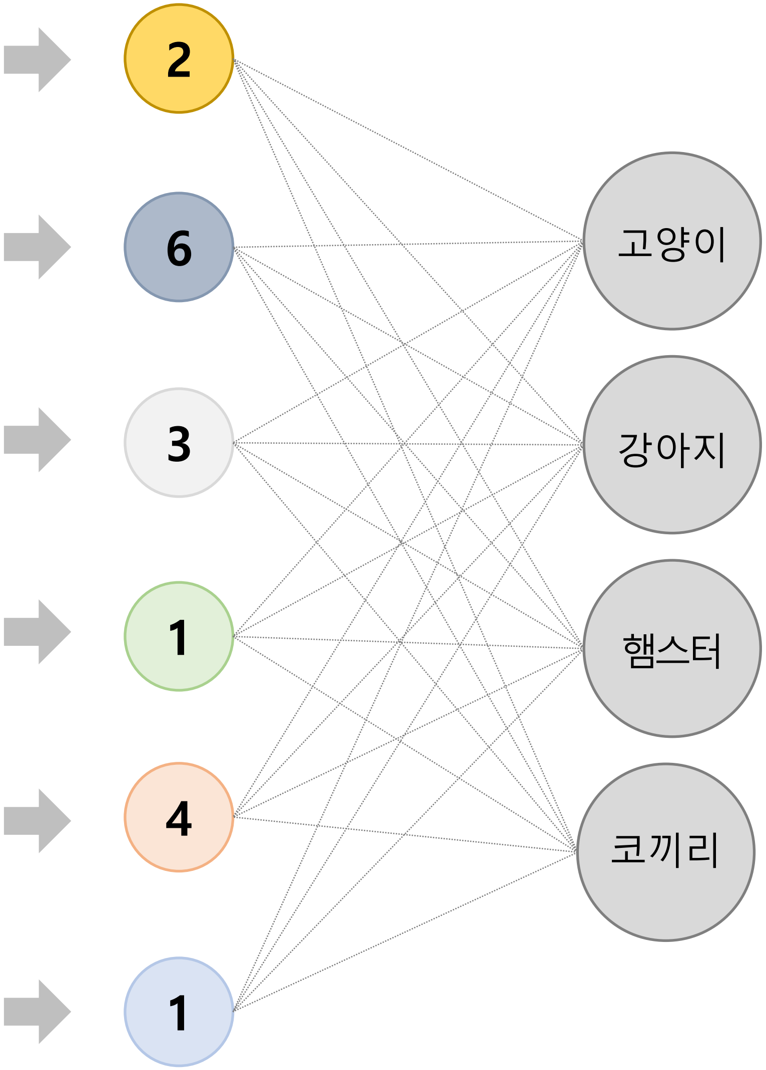
1	1	0
5	…	1
5	0	2

1	-1	0
-2	…	1
1	0	9

0	2	0
1	…	1
2	0	0

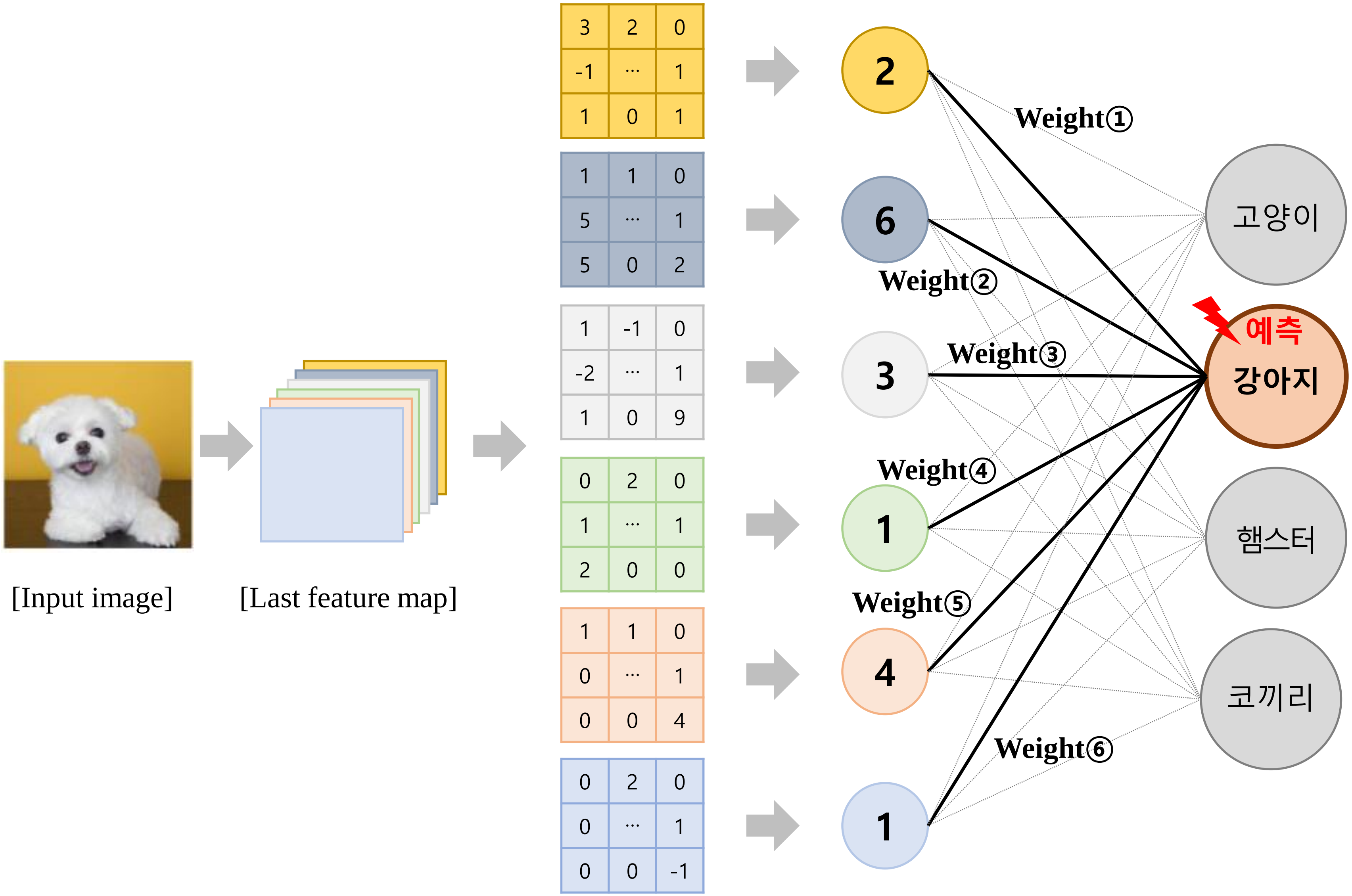
1	1	0
0	…	1
0	0	4

0	2	0
0	…	1
0	0	-1



Class Activation Map (CAM)

Global Average Pooling → 각 Feature별 평균 값을 구함



[Input image]

[Last feature map]

3	2	0
-1	…	1
1	0	1

1	1	0
5	…	1
5	0	2

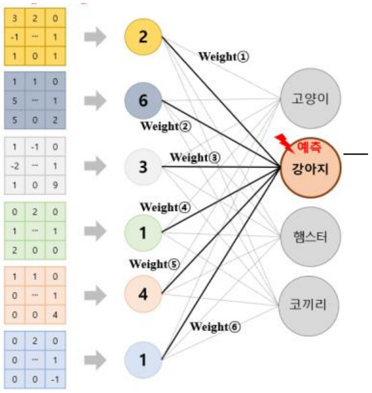
1	-1	0
-2	…	1
1	0	9

0	2	0
1	…	1
2	0	0

1	1	0
0	…	1
0	0	4

0	2	0
0	…	1
0	0	-1

Class Activation Map(CAM)



각 feature별 heatmap을 그림

3	2	0
-1	…	1
1	0	1

x Weight① =

1	1	0
5	…	1
5	0	2

x Weight② =

1	-1	0
-2	…	1
1	0	9

x Weight③ =

0	2	0
1	…	1
2	0	0

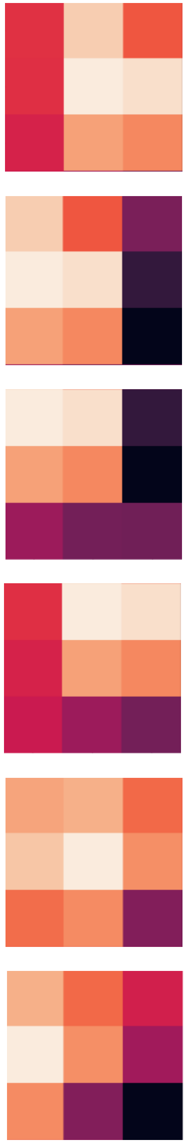
x Weight④ =

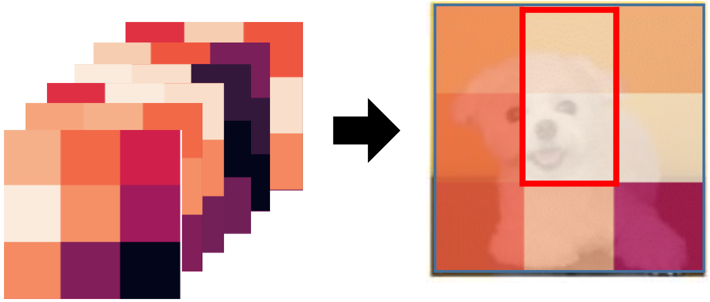
1	1	0
0	…	1
0	0	4

x Weight⑤ =

0	2	0
0	…	1
0	0	-1

x Weight⑥ =





동일 위치
Pixel별 합

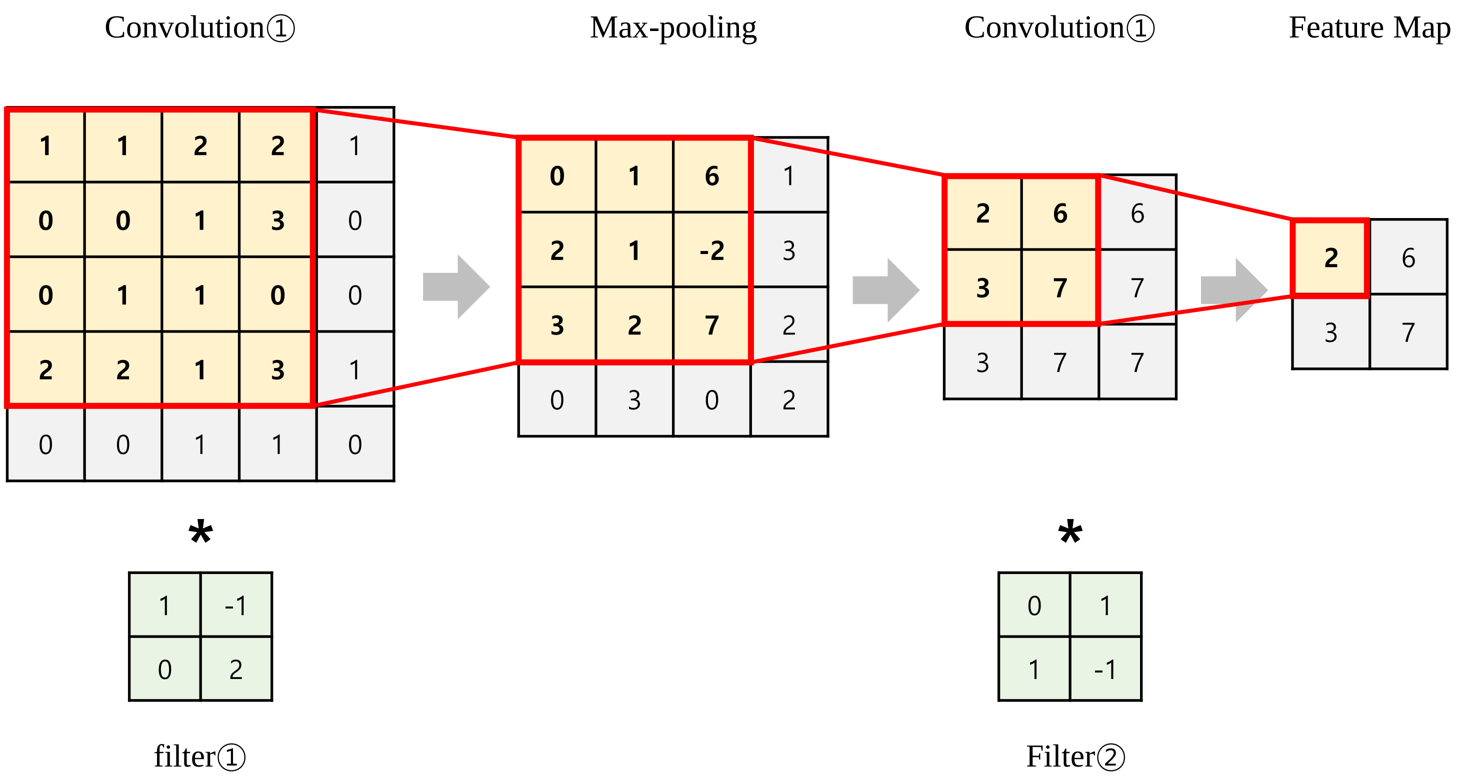
강아지 얼굴을
예측 원인으로 판단

Class Activation Map(CAM)

CAM으로 원인 분석이 가능한 이유

❖ Class Activation Map (CAM)에서 마지막 layer만으로도 원인 분석이 가능한 이유

- 마지막 feature map이 가진 정보량이 많기 때문에 원인 분석이 가능함
- 마지막 feature map 내 pixel 1개 값은 원본 이미지에서 많은 부분을 요약한 결과임



Convolution①

1	1	2	2	1
0	0	1	3	0
0	1	1	0	0
2	2	1	3	1
0	0	1	1	0

*

1	-1
0	2

filter①

Max-pooling

0	1	6	1
2	1	-2	3
3	2	7	2
0	3	0	2

Convolution①

2	6	6
3	7	7
3	7	7

*

0	1
1	-1

Filter②

Feature Map

2	6
3	7

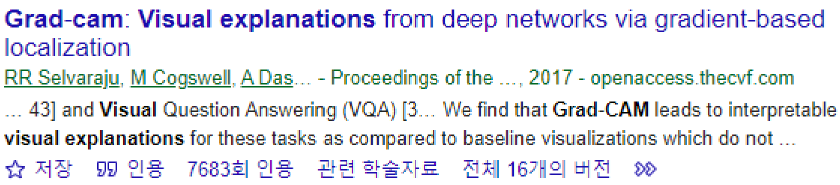
3. Gradient based CAM

Gradient based CAM

Grad-CAM 논문

❖ Grad-CAM (2017)

- Gradient를 활용해서 CAM 결과를 산출하는 방법론임
- 2017년도 ICCV (International Conference on Computer Vision)에서 소개됨
- 22년 2월 24일 기준 7,863회 인용됨

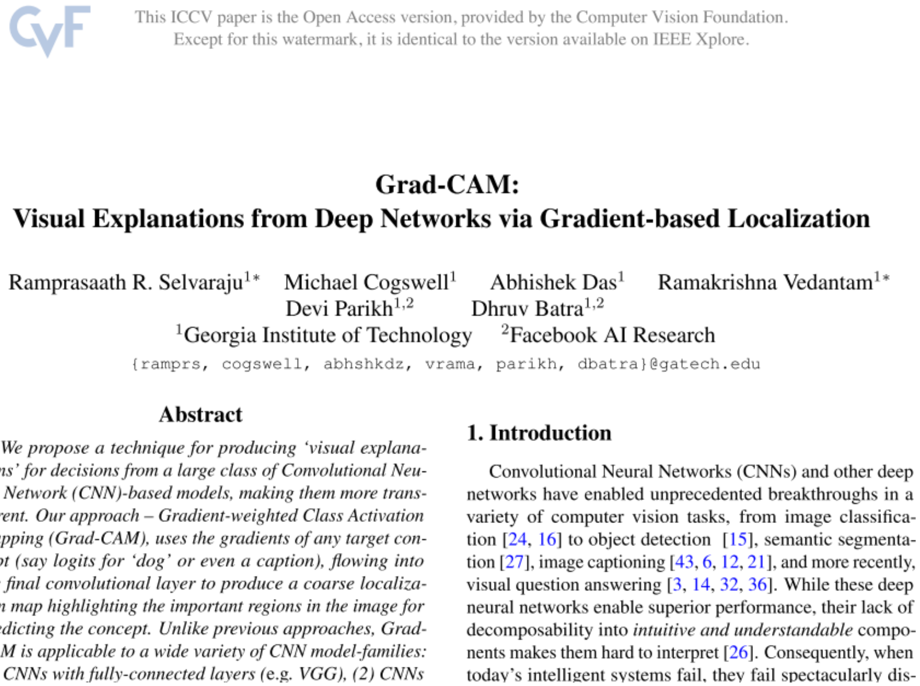


Grad-cam: Visual explanations from deep networks via gradient-based localization

RR Selvaraju, M Cogswell, A Das… - Proceedings of the …, 2017 - openaccess.thecvf.com

… 43] and **Visual** Question Answering (VQA) [3… We find that **Grad-CAM** leads to interpretable **visual explanations** for these tasks as compared to baseline visualizations which do not …

☆ 저장 99 인용 7683회 인용 관련 학술자료 전체 16개의 버전 »



**Grad-CAM:
Visual Explanations from Deep Networks via Gradient-based Localization**

Ramprasaath R. Selvaraju[1]* Michael Cogswell[1] Abhishek Das[1] Ramakrishna Vedantam[1]*
Devi Parikh[1,2] Dhruv Batra[1,2]
[1]Georgia Institute of Technology [2]Facebook AI Research
{ramprs, cogswell, abhshkdz, vrama, parikh, dbatra}@gatech.edu

Abstract

We propose a technique for producing 'visual explanations' for decisions from a large class of Convolutional Neural Network (CNN)-based models, making them more transparent. Our approach – Gradient-weighted Class Activation Mapping (Grad-CAM), uses the gradients of any target concept (say logits for 'dog' or even a caption), flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept. Unlike previous approaches, Grad-CAM is applicable to a wide variety of CNN model-families: (1) CNNs with fully-connected layers (e.g. VGG), (2) CNNs

1. Introduction

Convolutional Neural Networks (CNNs) and other deep networks have enabled unprecedented breakthroughs in a variety of computer vision tasks, from image classification [24, 16] to object detection [15], semantic segmentation [27], image captioning [43, 6, 12, 21], and more recently, visual question answering [3, 14, 32, 36]. While these deep neural networks enable superior performance, their lack of decomposability into *intuitive and understandable* components makes them hard to interpret [26]. Consequently, when today's intelligent systems fail, they fail spectacularly dis-

Reference: Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Gradient based CAM

Grad-CAM 전반적인 구조

❖ Grad-CAM(Class Activation Map) 구조

- CNN 구조 모델에서 gradient를 활용해 예측 결과에 대한 원인 해석을 진행
- GAP 구조가 필요 없기 때문에 모델 구조의 변경이 필요하지 않음

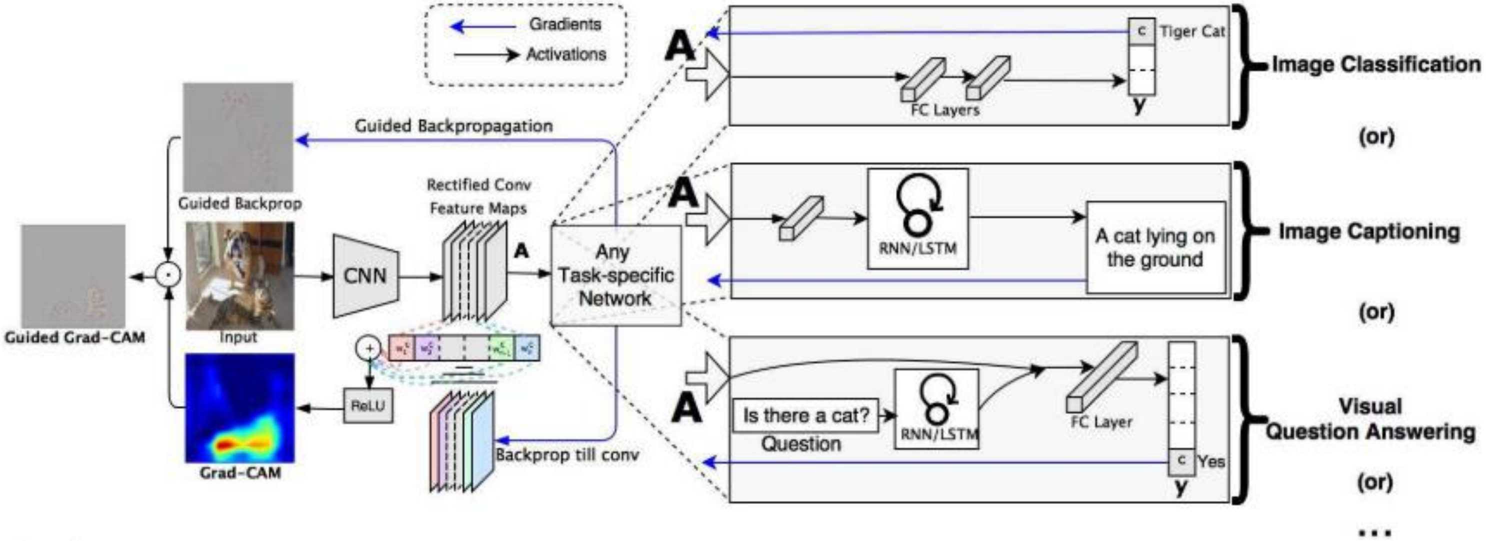


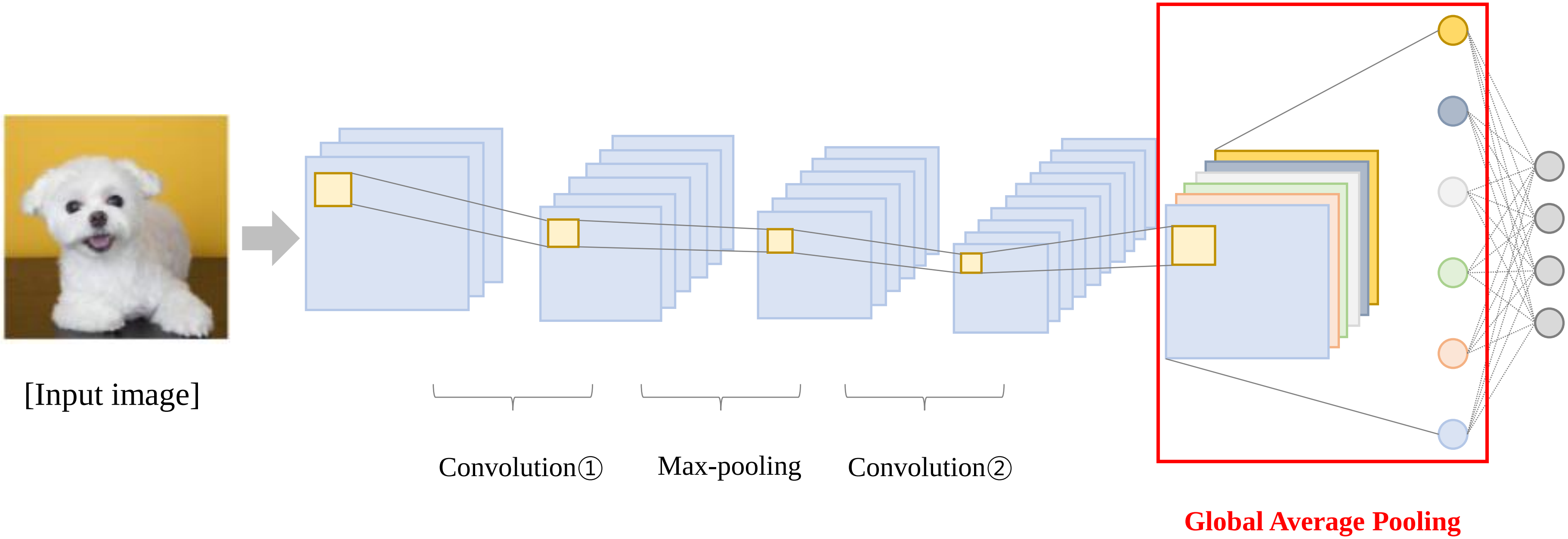
Figure 2: Grad-CAM overview: Given an image and a class of interest (*e.g.*, 'tiger cat' or any other type of differentiable output) as input, we forward propagate the image through the CNN part of the model and then through task-specific computations to obtain a raw score for the category. The gradients are set to zero for all classes except the desired class (tiger cat), which is set to 1. This signal is then backpropagated to the rectified convolutional feature maps of interest, which we combine to compute the coarse Grad-CAM localization (blue heatmap) which represents where the model has to look to make the particular decision. Finally, we pointwise multiply the heatmap with guided backpropagation to get Guided Grad-CAM visualizations which are both high-resolution and concept-specific.

Reference: Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization.
In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Gradient based CAM

CAM의 한계

- ❖ CNN + Class Activation Map(CAM) 구조
  - 마지막 Convolutional layer 뒤에 Global Average Pooling (GAP) 구조를 사용
  - CNN 구조에서 GAP 부분을 꼭 넣어줘야 하기 때문에 모델 구축에 제한이 발생함

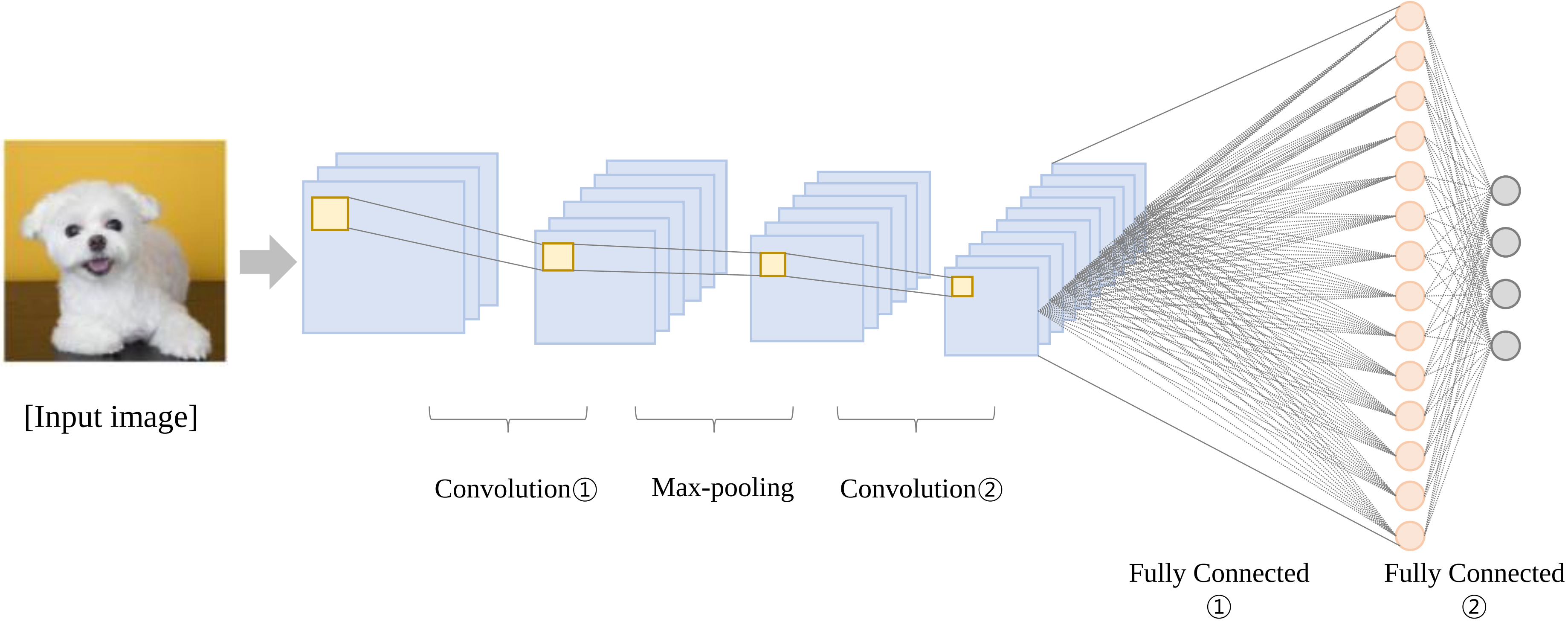


Gradient based CAM

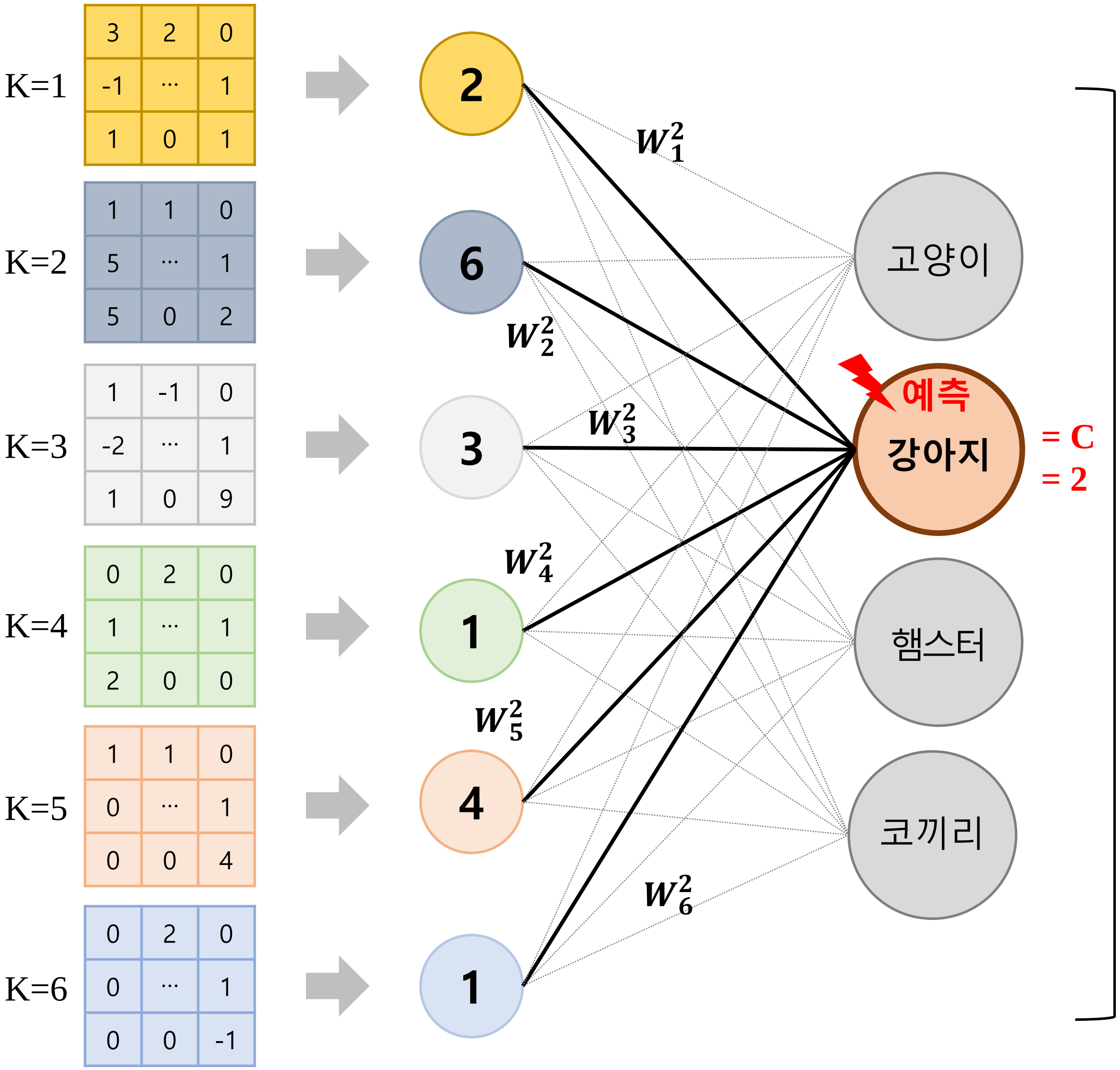
CAM의 한계

❖ CNN + Grad-CAM 구조

- Class Activation Map (CAM) 에서 마지막 convolution layer 뒤 GAP을 사용하지 않음
- CNN 기본 구조를 변형하지 않고 그대로 사용하기 때문에 기존 CNN 모델에 쉽게 적용할 수 있다는 장점이 존재함



Gradient based CAM



K=1

3	2	0
-1	…	1
1	0	1

K=2

1	1	0
5	…	1
5	0	2

K=3

1	-1	0
-2	…	1
1	0	9

K=4

0	2	0
1	…	1
2	0	0

K=5

1	1	0
0	…	1
0	0	4

K=6

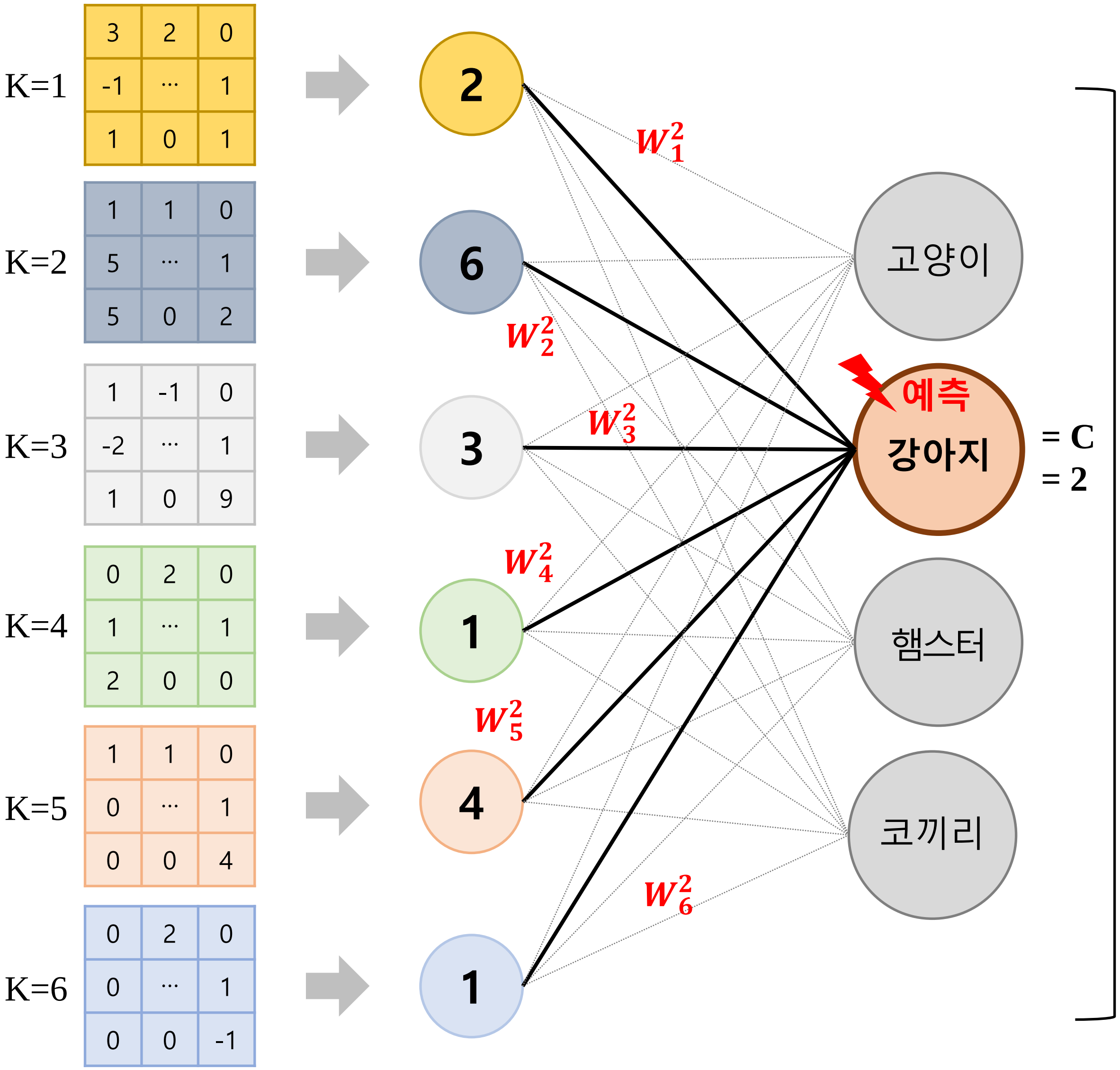
0	2	0
0	…	1
0	0	-1

<C class에 대한 CAM Score>

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

- C = 예측 Class
- W_k^c = C class 예측하는 k번째 Feature map에 대한 weight
- A^k= k번째 Feature map
- $A_{i,j}$= Feature map 내 i,j 위치 값
- Z = Feature map별 합

Gradient based CAM



K=1

3	2	0
-1	…	1
1	0	1

K=2

1	1	0
5	…	1
5	0	2

K=3

1	-1	0
-2	…	1
1	0	9

K=4

0	2	0
1	…	1
2	0	0

K=5

1	1	0
0	…	1
0	0	4

K=6

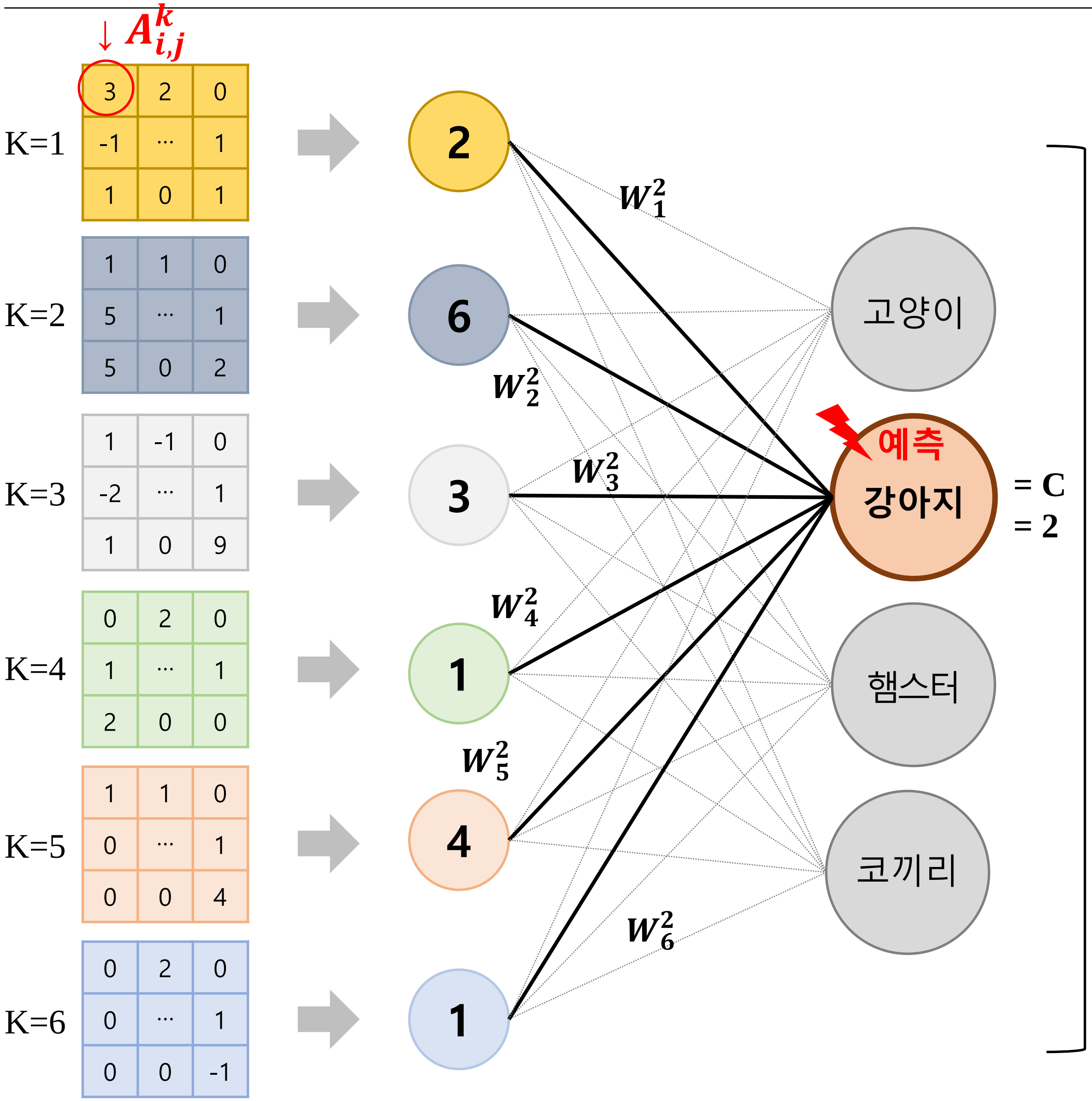
0	2	0
0	…	1
0	0	-1

<C class에 대한 CAM Score>

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

- C = 예측 Class
- **W_k^c = C class 예측하는 k번째 Feature map에 대한 weight**
- A^k= k번째 Feature map
- $A_{i,j}$= Feature map 내 i,j 위치 값
- Z = Feature map별 합

Gradient based CAM



K=1

3	2	0
-1	…	1
1	0	1

K=2

1	1	0
5	…	1
5	0	2

K=3

1	-1	0
-2	…	1
1	0	9

K=4

0	2	0
1	…	1
2	0	0

K=5

1	1	0
0	…	1
0	0	4

K=6

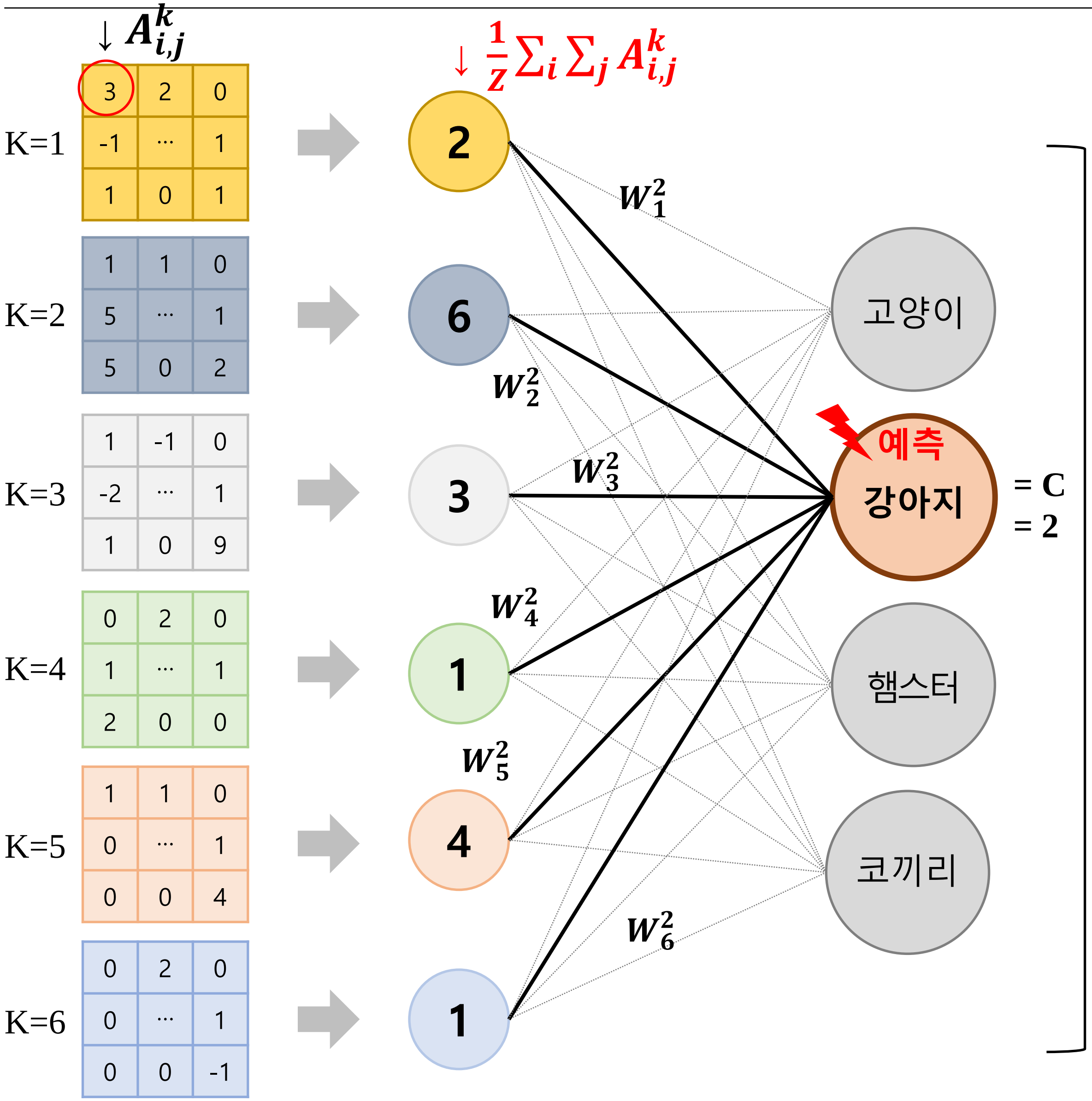
0	2	0
0	…	1
0	0	-1

<C class에 대한 CAM Score>

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

- C = 예측 Class
- W_k^c = C class 예측하는 k번째 Feature map에 대한 weight
- **A^k= k번째 Feature map**
- **$A_{i,j}$= Feature map 내 i,j 위치 값**
- Z = Feature map별 합

Gradient based CAM

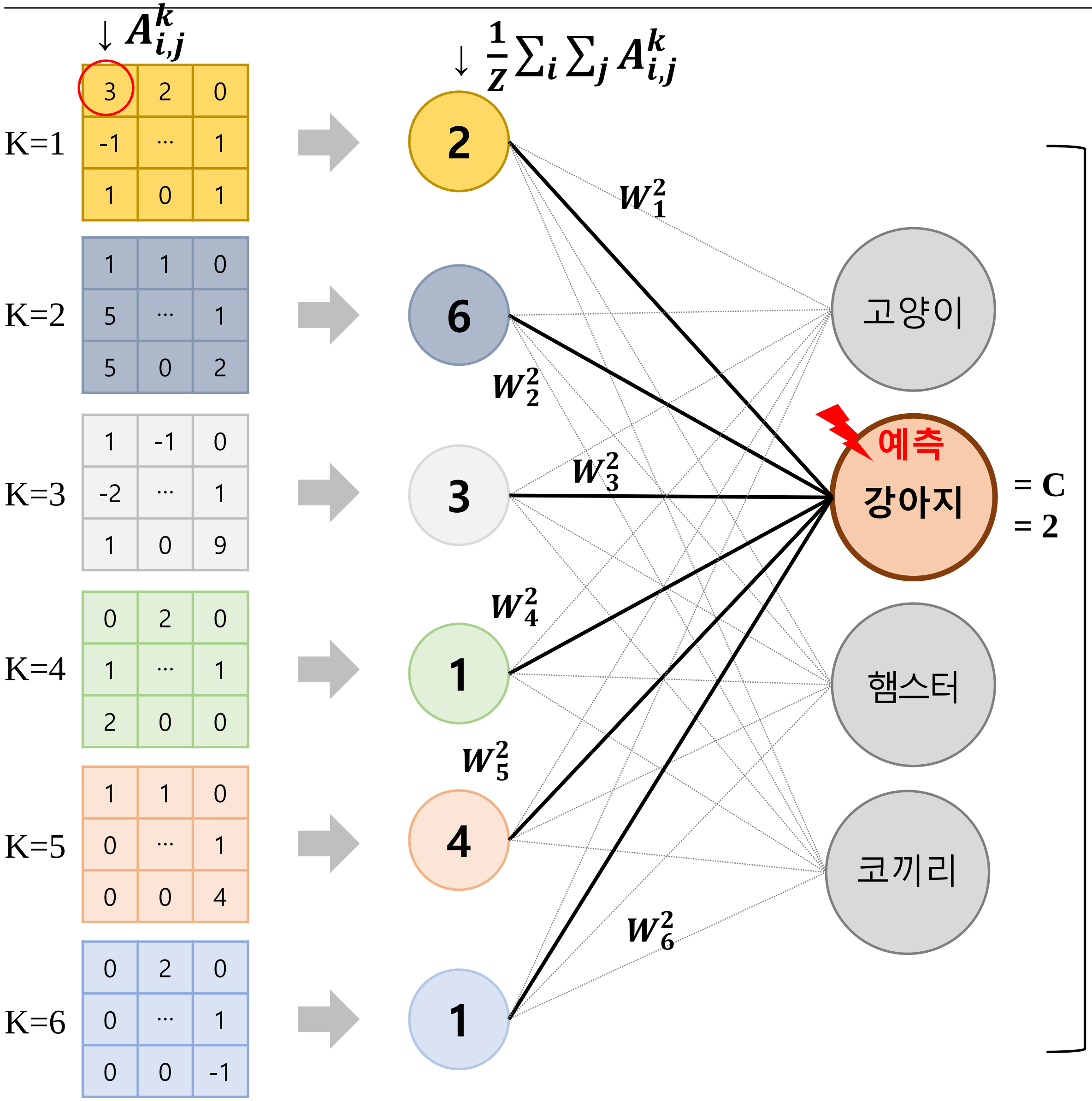


<C class에 대한 CAM Score>

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

- C = 예측 Class
- W_k^c = C class 예측하는 k번째 Feature map에 대한 weight
- **A^k= k번째 Feature map**
- **$A_{i,j}$= Feature map 내 i,j 위치 값**
- **Z = Feature map별 합**

Gradient based CAM



K=1

3	2	0
-1	…	1
1	0	1

K=2

1	1	0
5	…	1
5	0	2

K=3

1	-1	0
-2	…	1
1	0	9

K=4

0	2	0
1	…	1
2	0	0

K=5

1	1	0
0	…	1
0	0	4

K=6

0	2	0
0	…	1
0	0	-1

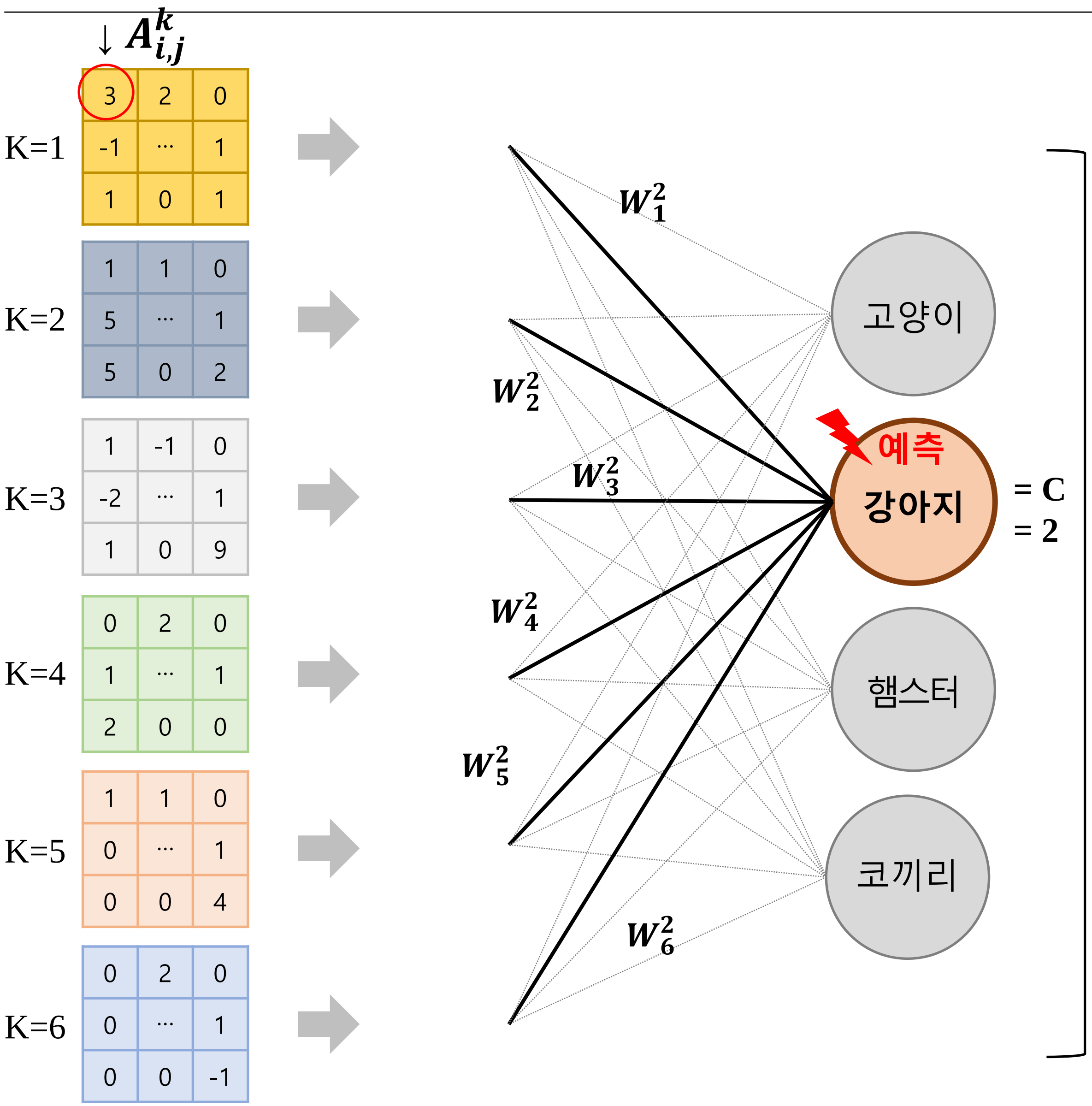
여기서 GAP을 사용하지 않으면?

<C class에 대한 CAM Score>

$$S^c = \sum_k W^c_k \frac{1}{Z} \sum_i \sum_j A^k_{i,j}$$

- C = 예측 Class
- W^c_k = C class 예측하는 k번째 Feature map에 대한 weight
- A^k= k번째 Feature map
- $A_{i,j}$= Feature map 내 i,j 위치 값
- Z = Feature map별 합

Gradient based CAM



K=1

3	2	0
-1	···	1
1	0	1

K=2

1	1	0
5	···	1
5	0	2

K=3

1	-1	0
-2	···	1
1	0	9

K=4

0	2	0
1	···	1
2	0	0

K=5

1	1	0
0	···	1
0	0	4

K=6

0	2	0
0	···	1
0	0	-1

여기서 GAP을 사용하지 않으면?

→ Feature map별 weight도 사용하지 못함

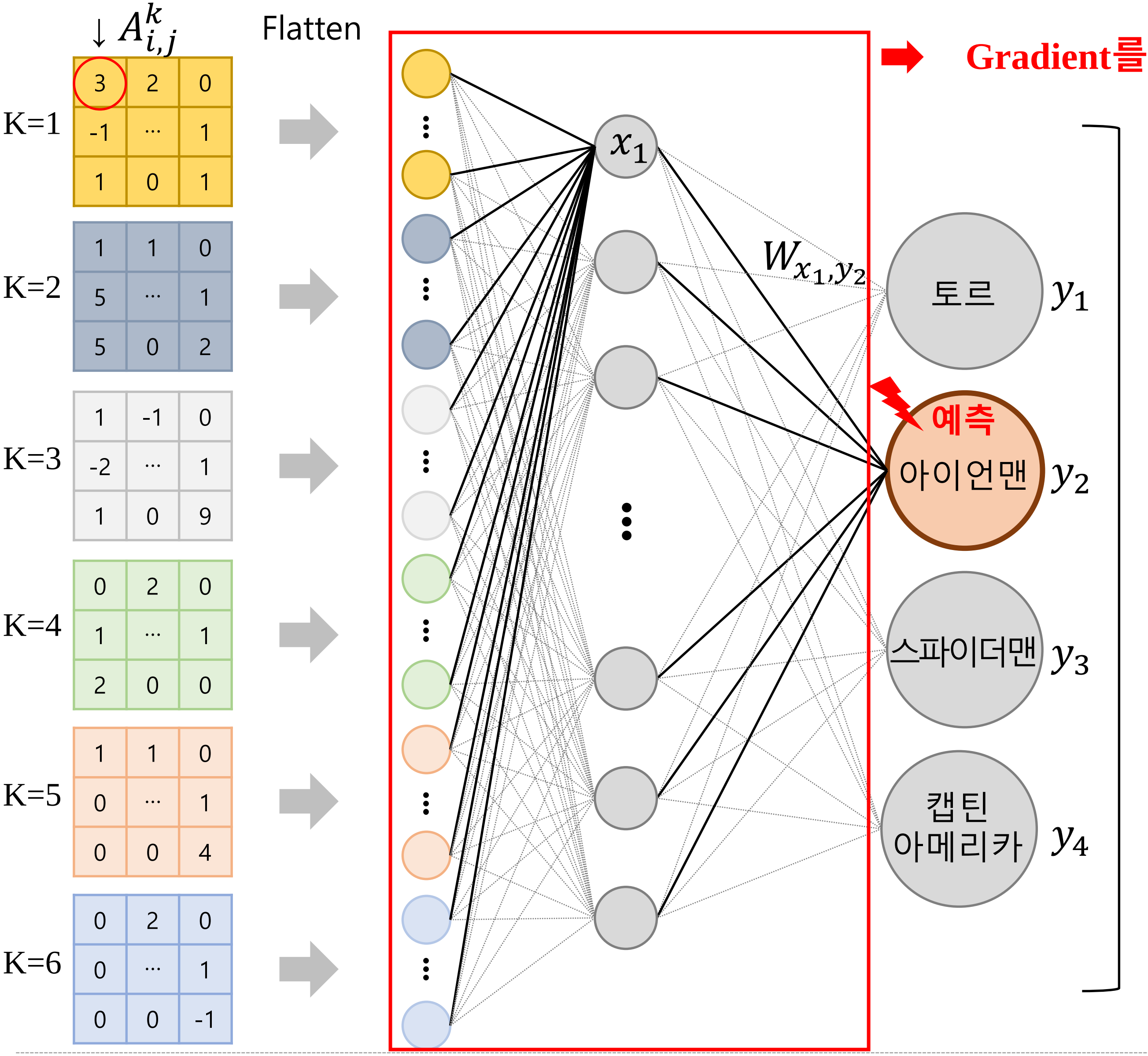
→ Weight를 정의하는 방식을 바꾸자

<C class에 대한 CAM Score>

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

- C = 예측 Class
- W_k^c = C class 예측하는 k번째 Feature map에 대한 weight
- A^k= k번째 Feature map
- $A_{i,j}$= Feature map 내 i,j 위치 값
- Z = Feature map별 합

Gradient based CAM



K=1

3	2	0
-1	···	1
1	0	1

K=2

1	1	0
5	···	1
5	0	2

K=3

1	-1	0
-2	···	1
1	0	9

K=4

0	2	0
1	···	1
2	0	0

K=5

1	1	0
0	···	1
0	0	4

K=6

0	2	0
0	···	1
0	0	-1

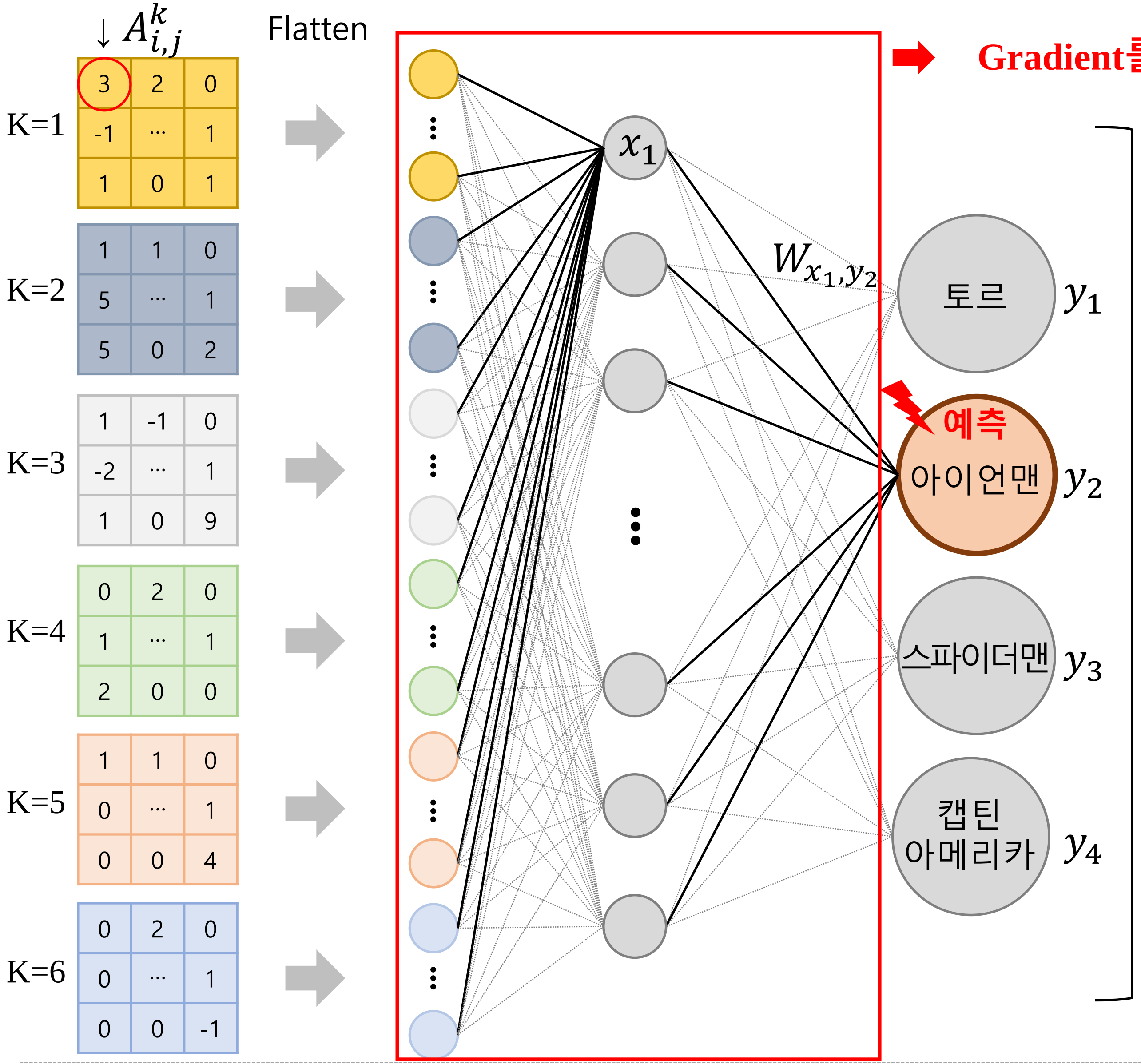
➡ **Gradient를 통해 weight를 구하자**

$$\sum_i \sum_j \frac{\partial y_c}{\partial A^k_{i,j}}$$

$$y_c = Wx + b$$

$$y_2 = W_{x_1,y_2} x_1 + b$$

Gradient based CAM



K=1

3	2	0
-1	…	1
1	0	1

K=2

1	1	0
5	…	1
5	0	2

K=3

1	-1	0
-2	…	1
1	0	9

K=4

0	2	0
1	…	1
2	0	0

K=5

1	1	0
0	…	1
0	0	4

K=6

0	2	0
0	…	1
0	0	-1

➡ **Gradient를 통해 weight를 구하자**

<C class에 대한 CAM Score>

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

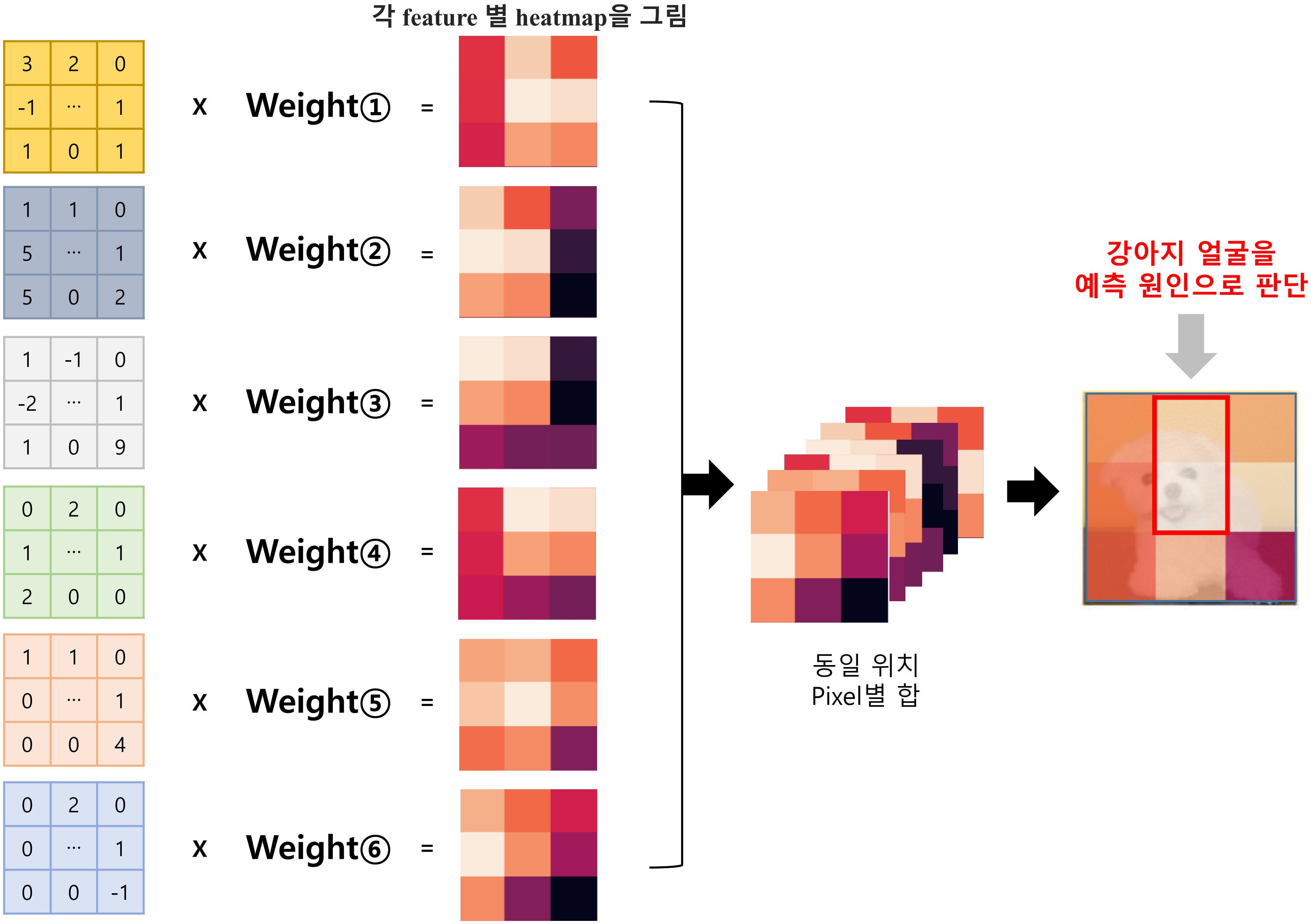
⬇

< C class에 대한 Grad_CAM Score >

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

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

Gradient based CAM



각 feature 별 heatmap을 그림

3	2	0
-1	…	1
1	0	1

x Weight① =

1	1	0
5	…	1
5	0	2

x Weight② =

1	-1	0
-2	…	1
1	0	9

x Weight③ =

0	2	0
1	…	1
2	0	0

x Weight④ =

1	1	0
0	…	1
0	0	4

x Weight⑤ =

0	2	0
0	…	1
0	0	-1

x Weight⑥ =

동일 위치
Pixel별 합

강아지 얼굴을
예측 원인으로 판단

Gradient based CAM

각 feature 별 heatmap을 그림

3	2	0
-1	···	1
1	0	1

x $\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A^1_{i,j}}$ =

1	1	0
5	···	1
5	0	2

x $\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A^2_{i,j}}$ =

1	-1	0
-2	···	1
1	0	9

x $\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A^3_{i,j}}$ =

0	2	0
1	···	1
2	0	0

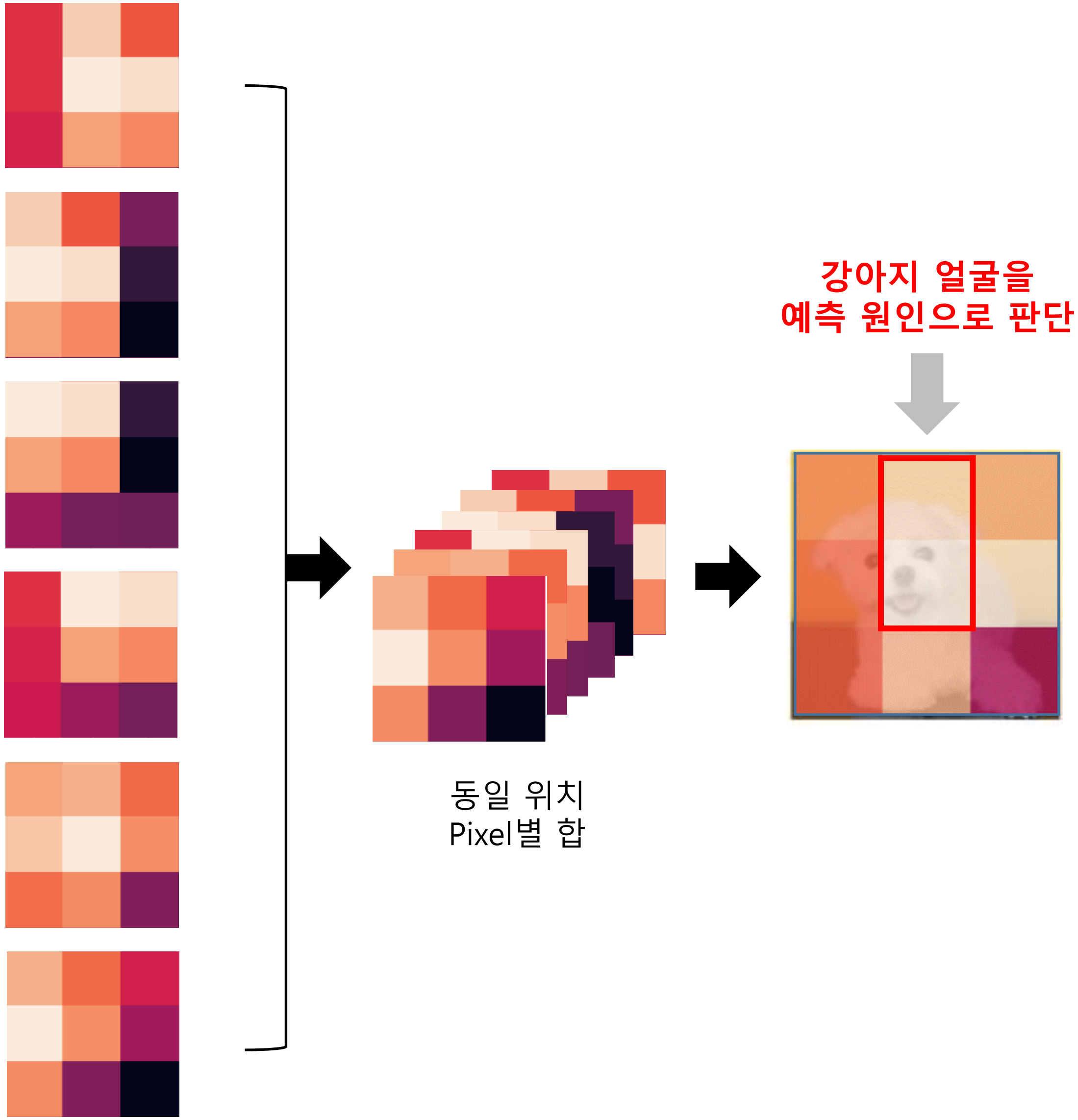
x $\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A^4_{i,j}}$ =

1	1	0
0	···	1
0	0	4

x $\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A^5_{i,j}}$ =

0	2	0
0	···	1
0	0	-1

x $\frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A^6_{i,j}}$ =



동일 위치
Pixel별 합

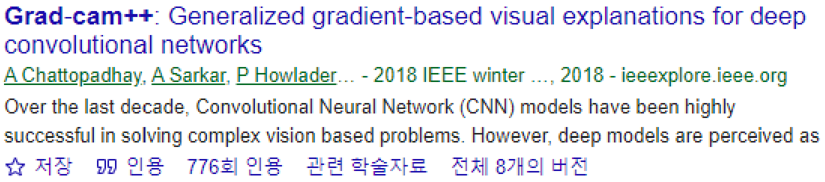
강아지 얼굴을
예측 원인으로 판단

Gradient based CAM

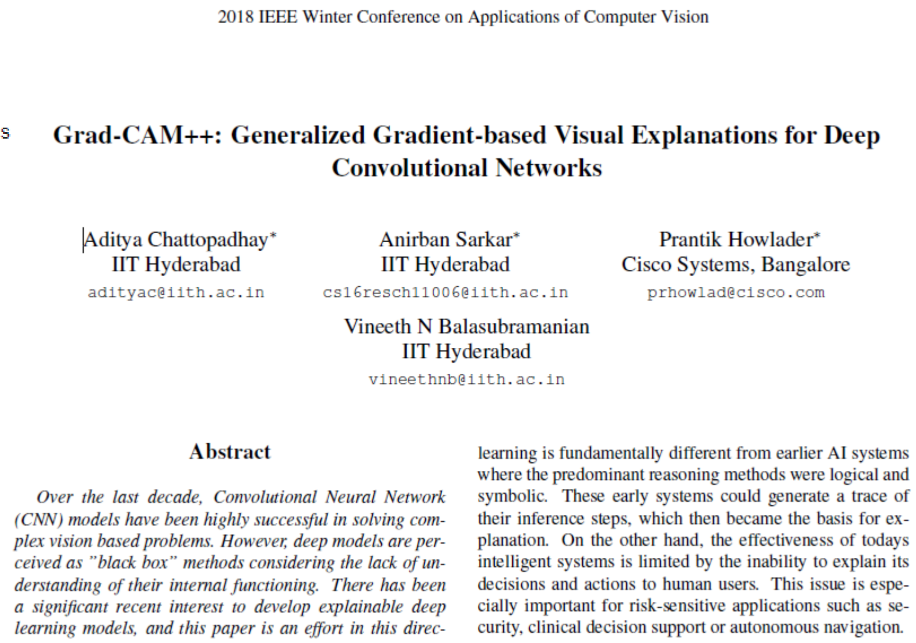
Grad-CAM++ 논문

❖ Grad-CAM++ (2018)

- Grad-CAM을 일반화한 버전으로 Grad-CAM의 한계점을 보완함
- 2018년도 WACV(Winter conference on Application of Computer Vision)에서 발표됨
- 22년 2월 24일 기준 776회 인용됨



Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks
A Chattopadhay, A Sarkar, P Howlader… - 2018 IEEE winter …, 2018 - ieeexplore.ieee.org
Over the last decade, Convolutional Neural Network (CNN) models have been highly successful in solving complex vision based problems. However, deep models are perceived as
☆ 저장 99 인용 776회 인용 관련 학술자료 전체 8개의 버전



2018 IEEE Winter Conference on Applications of Computer Vision

Grad-CAM++: Generalized Gradient-based Visual Explanations for Deep Convolutional Networks

Aditya Chattopadhay* IIT Hyderabad adityac@iith.ac.in
Anirban Sarkar* IIT Hyderabad cs16resch11006@iith.ac.in
Prantik Howlader* Cisco Systems, Bangalore prhowlad@cisco.com
Vineeth N Balasubramanian IIT Hyderabad vineethnb@iith.ac.in

Abstract

Over the last decade, Convolutional Neural Network (CNN) models have been highly successful in solving complex vision based problems. However, deep models are perceived as "black box" methods considering the lack of understanding of their internal functioning. There has been a significant recent interest to develop explainable deep learning models, and this paper is an effort in this direc-

learning is fundamentally different from earlier AI systems where the predominant reasoning methods were logical and symbolic. These early systems could generate a trace of their inference steps, which then became the basis for explanation. On the other hand, the effectiveness of todays intelligent systems is limited by the inability to explain its decisions and actions to human users. This issue is especially important for risk-sensitive applications such as security, clinical decision support or autonomous navigation.

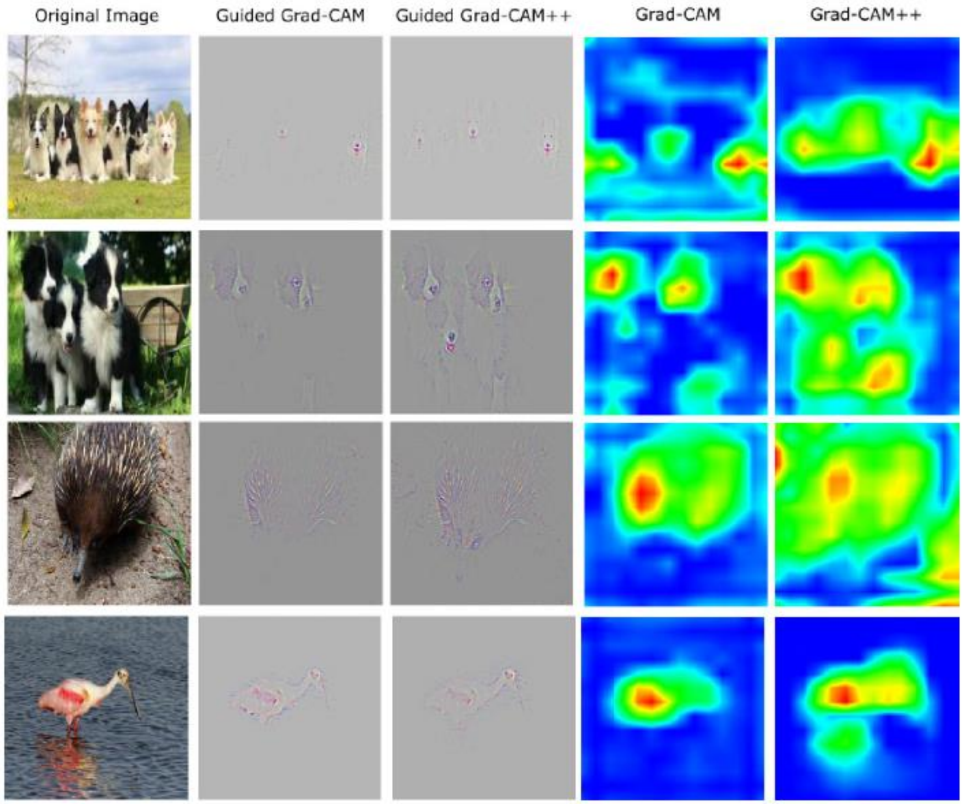
Reference: Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE.

Gradient based CAM

Grad-CAM의 한계점

❖ Grad-CAM의 한계점

- Grad-CAM은 이미지 내 다중 객체가 있을 때, 이를 정확히 포착하지 못함 (1,2번)
- 단일 객체인 경우에도 넓은 영역일 때는 중간에 끊기는 경우가 존재함 (3,4번)



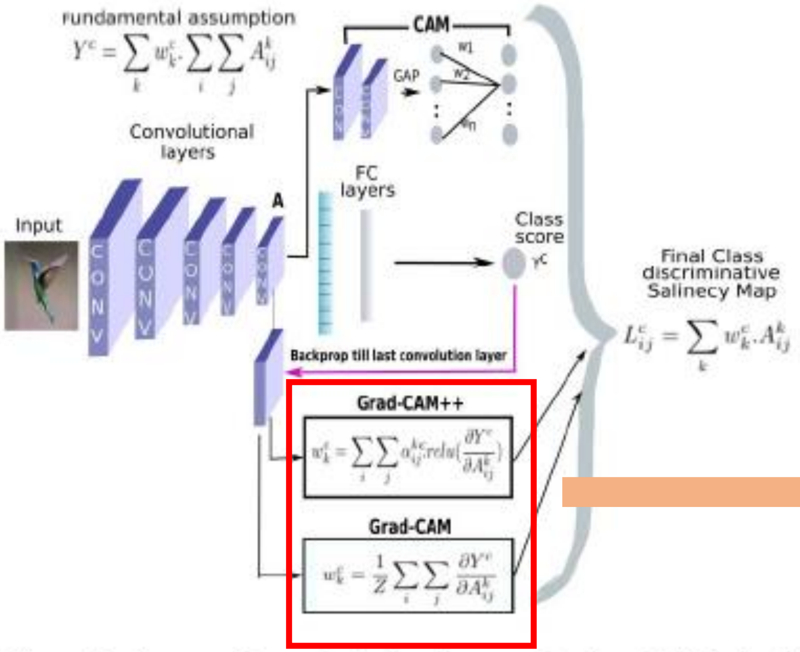
Reference: Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE.

Gradient based CAM

Grad-CAM++ 구조

❖ Grad-CAM++ 구조 (2018)

- Grad-CAM과 같은 구조를 지니지만, weight를 구하는 부분이 다름
- Grad-CAM을 일반화한 버전으로 gradient의 weighted average를 normalized term으로 활용



fundamental assumption
$Y^c = \sum_k w_k^c . \sum_i \sum_j A_{ij}^k$
CAM
w1
w2
wn
GAP
Convolutional layers
FC layers
Input
CONV
CONV
CONV
CONV
CONV
A
Class score
Y^c
Final Class discriminative Salinecy Map
$L_{ij}^c = \sum_k w_k^c . A_{ij}^k$
Backprop till last convolution layer
Grad-CAM++
$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc} relu(\frac{\partial Y^c}{\partial A_{ij}^k})$
Grad-CAM
$w_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial Y^c}{\partial A_{ij}^k}$

Figure 3. An overview of all the three methods – CAM, Grad-CAM, Grad-CAM++ – with their respective computation expressions.

<Weight 계산식>

<Grad-CAM++>

$$W_k^c = \sum_i \sum_j \alpha_{i,j}^{k,c} \, relu \frac{\partial y_c}{\partial A_{i,j}^k}$$

<Grad-CAM>

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

Reference: Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE.

Gradient based CAM

Grad-CAM++ 구조

❖ Grad-CAM++ 구조 (2018)

- Grad-CAM과 같은 구조를 지니지만, weight를 구하는 부분이 다름
- Grad-CAM을 일반화한 버전으로 gradient의 weighted average를 normalized term으로 활용
- 수식의 전개 과정은 아래 reference에 자세히 적혀 있음

<Weight 계산식>

<Grad-CAM++>

$$W_k^c = \sum_i \sum_j \boldsymbol{\alpha_{i,j}^{k,c}} \, \boldsymbol{relu} \frac{\partial y_c}{\partial A_{i,j}^k}$$

<Grad-CAM>

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

$$\alpha_{i,j}^{k,c} = \frac{\frac{\partial^2 y_c}{(\partial A_{i,j}^k)^2}}{2\frac{\partial^2 y_c}{(\partial A_{i,j}^k)^2} + \sum_a \sum_b A_{a,b}^k \{\frac{\partial^3 y_c}{(\partial A_{i,j}^k)^3}\}}$$

만약 $\alpha_{i,j}^{k,c} = 1/Z$ 이면,
Grad-CAM과 같은 수식임

Reference: Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE.

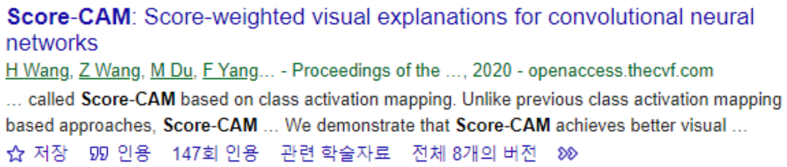
5. Score-CAM

Score-CAM

Score-CAM 논문

❖ Score-CAM (2020)

- Gradient based CAM의 문제점을 보완해 visualization 성능을 더 높인 연구임
- 2020년도 CVPR (Conference on Computer Vision and Pattern Recognition workshop)에서 발표됨
- 22년 2월 24일 기준 147회 인용됨

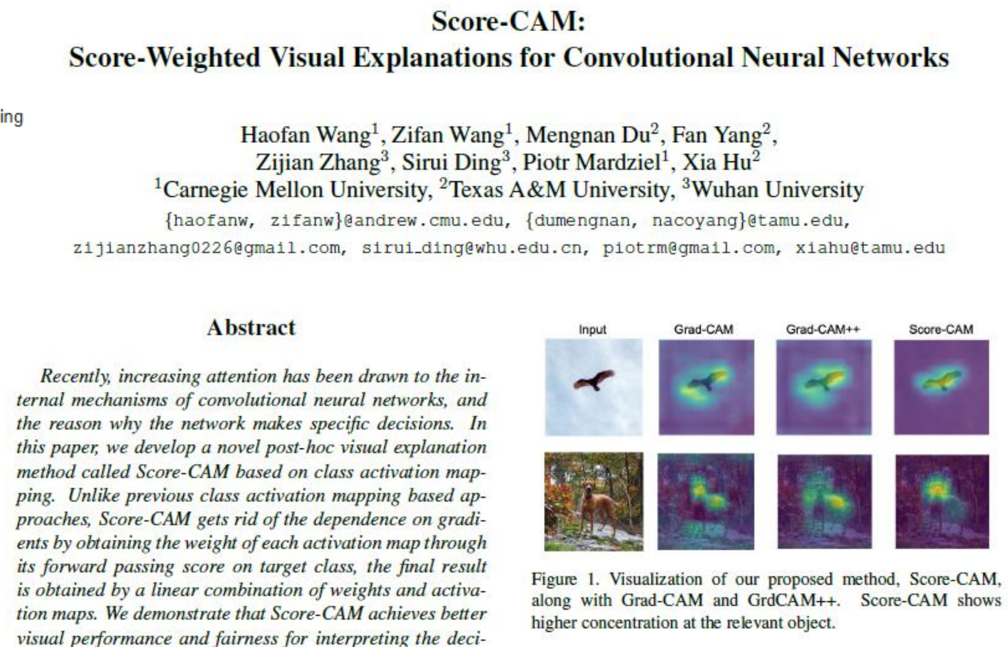


Score-CAM: Score-weighted visual explanations for convolutional neural networks

H Wang, Z Wang, M Du, F Yang... - Proceedings of the ..., 2020 - openaccess.thecvf.com

... called **Score-CAM** based on class activation mapping. Unlike previous class activation mapping based approaches, **Score-CAM** ... We demonstrate that **Score-CAM** achieves better visual ...

☆ 저장 99 인용 147회 인용 관련 학술자료 전체 8개의 버전 ≫



Score-CAM:
Score-Weighted Visual Explanations for Convolutional Neural Networks

Haofan Wang[1], Zifan Wang[1], Mengnan Du[2], Fan Yang[2], Zijian Zhang[3], Sirui Ding[3], Piotr Mardziel[1], Xia Hu[2]
[1]Carnegie Mellon University, [2]Texas A&M University, [3]Wuhan University
{haofanw, zifanw}@andrew.cmu.edu, {dumengnan, nacoyang}@tamu.edu, zijianzhang0226@gmail.com, sirui_ding@whu.edu.cn, piotrm@gmail.com, xiahu@tamu.edu

Abstract

Recently, increasing attention has been drawn to the internal mechanisms of convolutional neural networks, and the reason why the network makes specific decisions. In this paper, we develop a novel post-hoc visual explanation method called Score-CAM based on class activation mapping. Unlike previous class activation mapping based approaches, Score-CAM gets rid of the dependence on gradients by obtaining the weight of each activation map through its forward passing score on target class, the final result is obtained by a linear combination of weights and activation maps. We demonstrate that Score-CAM achieves better visual performance and fairness for interpreting the deci-

Figure 1. Visualization of our proposed method, Score-CAM, along with Grad-CAM and GrdCAM++. Score-CAM shows higher concentration at the relevant object.

Reference: Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

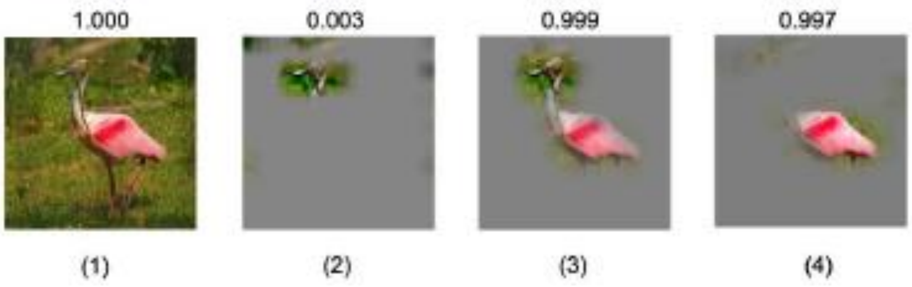
Score-CAM

Gradient based CAM 문제점

❖ 기존 Gradient 기반 CAM의 문제점

1) Gradient issue: gradient 계산시 noise가 존재하며, 활성화 함수 (ReLU, sigmoid) 사용으로 gradient가 0이 되는 saturation 현상이 발생함

2) False confidence: 활성화 맵 (Activation map) 부분의 weight가 더 크지만 target score는 오히려 작아지는 반대 현상이 일어나는 경우가 존재함



Weight:	Input image	0.035	0.027	0.021
Target score:	1.000	0.003	0.999	0.997

Reference: Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

Score-CAM

Score-CAM 전반적인 구조

❖ Score-CAM 구조 (2020)

- Gradient를 사용하지 않고 Channel-wise Increase of Confidence (CIC)로 가중치 구함
- CIC: Baseline 이미지와 channel별 feature 값의 차이를 통해 중요도를 나타냄

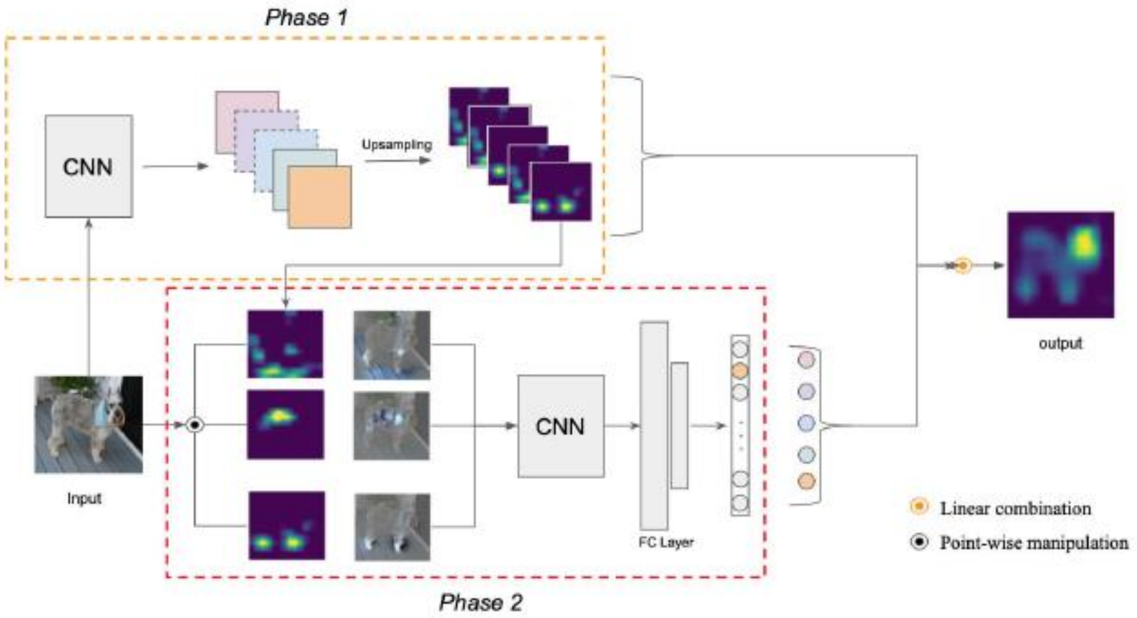


Figure 3. Pipeline of our proposed Score-CAM. Activation maps are first extracted in Phase 1. Each activation then works as a mask on original image, and obtain its forward-passing score on the target class. Phase 2 repeats for N times where N is the number of activation maps. Finally, the result can be generated by linear combination of score-based weights and activation maps. Phase 1 and Phase 2 shares a same CNN module as feature extractor.

Reference: Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

Score-CAM

$\alpha_k^c = W_k^c$

(논문) (세미나)

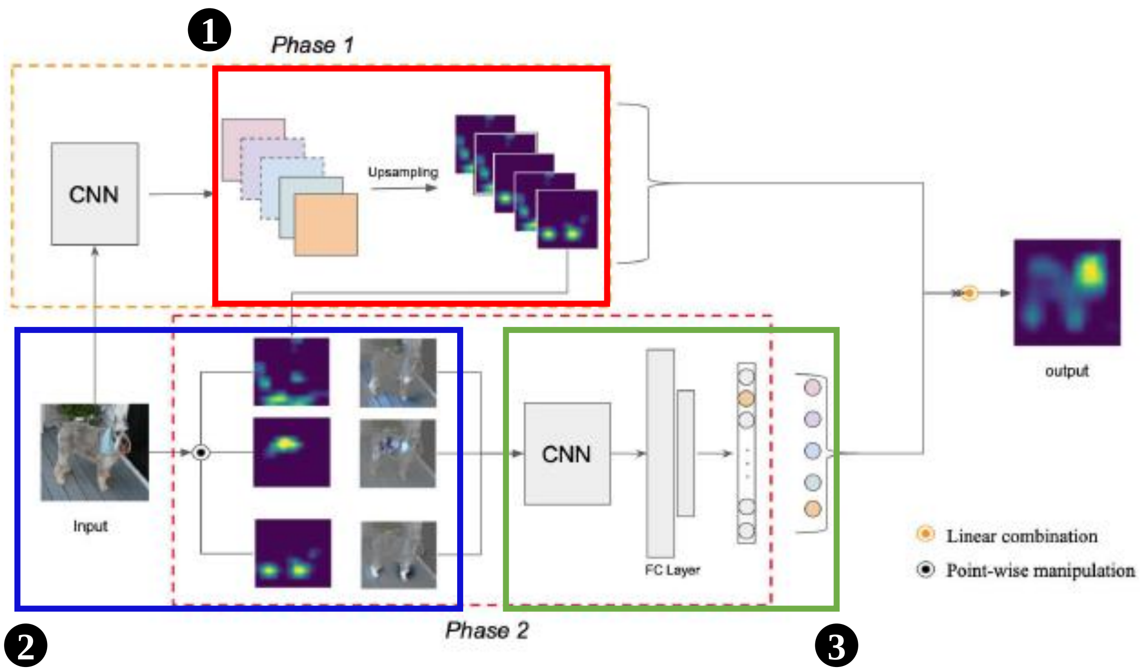
Score-CAM weight

❖ Score-CAM weight 계산 방식

① 마지막 feature map을 upsampling한 후, normalize를 해서 activation map (H_l^k)을 산출함

② 입력 이미지와 pixel (point)별 곱셈 연산을 진행해 ($X \cdot H_l^k$) 산출함

③ ($X \cdot H_l^k$)를 CNN 모델로 연산 후, baseline image (여기서는 0) 를 빼서 weight (W_k^c) 산출함



❸ ❷

$$W_k^c = C(A_l^k) = f(X \cdot H_l^k) - f(X_b)$$

❶

where, $H_l^k = s(Up(A_l^k))$

$f(X_b) = 0$ (black image)

- A_l^k = l번째 k채널의 feature map
- Up = upsampling
- S = normalized term

$$L_{Score-CAM}^c = ReLU \sum_k W_k^c A^k$$

Reference: Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

Score-CAM

Score-CAM 전반적인 논문

❖ Score-CAM 구조 (2020)

- Score-CAM에서는 Gradient를 사용하지 않고 채널별 중요도 차이로 가중치를 계산함

<Grad-CAM>

$$L^c_{Grad-CAM} = ReLU \sum_k W^c_k A^k$$

$$\boldsymbol{W^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y_c}{\partial A^k_{i,j}}}$$

<Score CAM>

$$L^c_{Score-CAM} = ReLU \sum_k W^c_k A^k$$

$$\boldsymbol{W^c_k = C\left(A^k_l\right) = f\left(X \cdot H^k_l\right) - f\left(X_b\right)}$$

where, $H^k_l = s\left(Up\left(A^k_l\right)\right)$

$f\left(X_b\right)$ = 0 (black image)

- A^k_l = l번 째 k채널의 $feature\ map$
- $Up = upsampling$
- $S = normalized\ term$

Reference: Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

Score-CAM

❖ Score-CAM 결과

- 단일 객체에 대해서 gradient based CAM 대비 더 좋은 성능을 보임
- 복수 객체에 대해서도 객체가 존재하는 부분을 더 잘 포착함을 확인

<단일 객체>

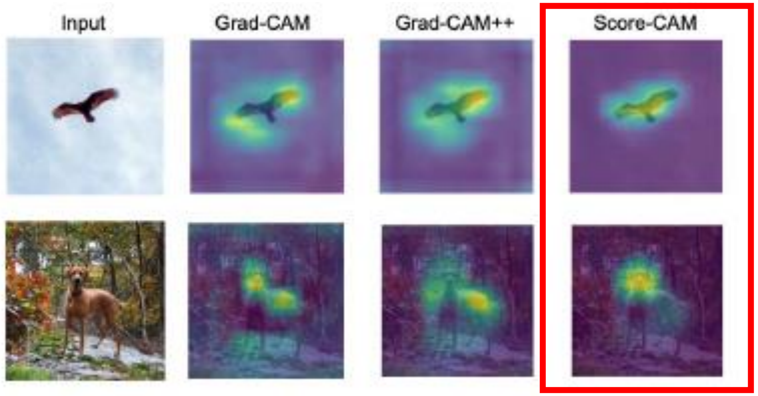


Figure 1. Visualization of our proposed method, Score-CAM, along with Grad-CAM and GrdCAM++. Score-CAM shows higher concentration at the relevant object.

<복수 객체>

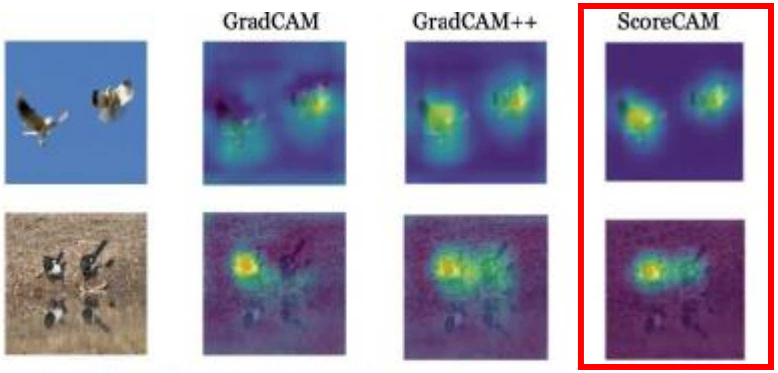


Figure 7. Results on multiple objects. As shown in this example, Grad-CAM only tends to focus on one object, while Grad-CAM++ can highlight all objects. Score-CAM further improves the quality of finding all evidences.

Reference: Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).

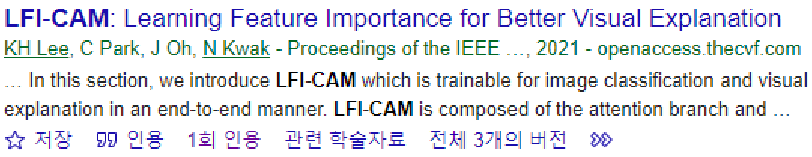
6. LFI-CAM

LFI-CAM

LFI-CAM 논문

❖ LFI-CAM (2021)

- Attention map을 활용해서 CAM의 가중치를 구함, Score-CAM 대비 연산 속도 빠름
- 2021년도 ICCV (International Conference on Computer Vision)에서 발표됨
- 22년 2월 24일 기준 1회 인용됨

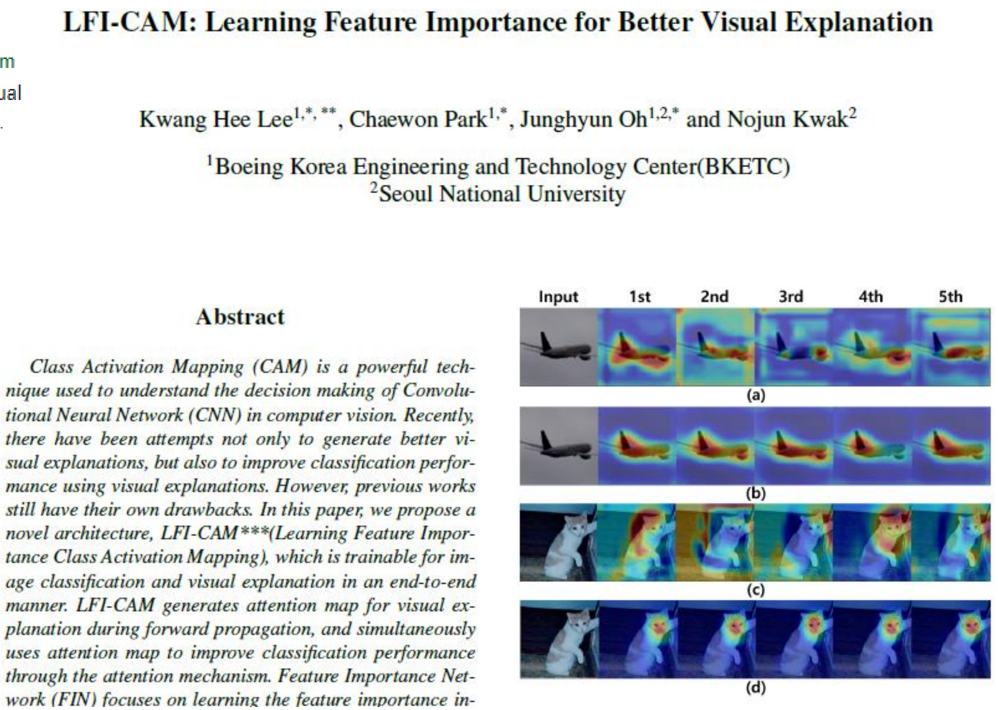


LFI-CAM: Learning Feature Importance for Better Visual Explanation

KH Lee, C Park, J Oh, N Kwak - Proceedings of the IEEE ..., 2021 - openaccess.thecvf.com

... In this section, we introduce **LFI-CAM** which is trainable for image classification and visual explanation in an end-to-end manner. **LFI-CAM** is composed of the attention branch and ...

☆ 저장 99 인용 1회 인용 관련 학술자료 전체 3개의 버전 »



LFI-CAM: Learning Feature Importance for Better Visual Explanation

Kwang Hee Lee[1,*,**], Chaewon Park[1,*], Junghyun Oh[1,2,*] and Nojun Kwak[2]

[1]Boeing Korea Engineering and Technology Center(BKETC)
[2]Seoul National University

Abstract

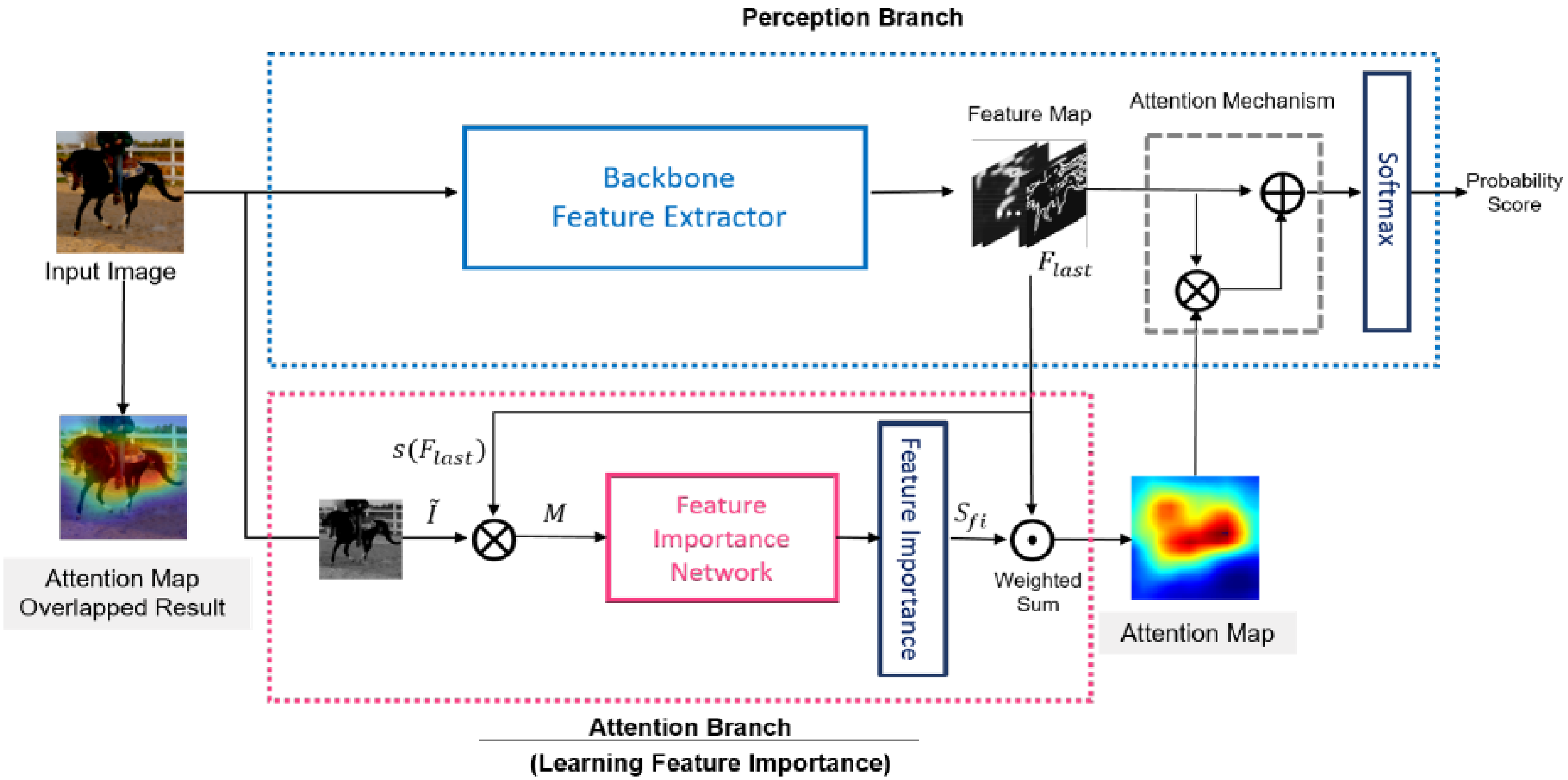
*Class Activation Mapping (CAM) is a powerful technique used to understand the decision making of Convolutional Neural Network (CNN) in computer vision. Recently, there have been attempts not only to generate better visual explanations, but also to improve classification performance using visual explanations. However, previous works still have their own drawbacks. In this paper, we propose a novel architecture, LFI-CAM***(Learning Feature Importance Class Activation Mapping), which is trainable for image classification and visual explanation in an end-to-end manner. LFI-CAM generates attention map for visual explanation during forward propagation, and simultaneously uses attention map to improve classification performance through the attention mechanism. Feature Importance Network (FIN) focuses on learning the feature importance in-*

Reference: Lee, K. H., Park, C., Oh, J., & Kwak, N. (2021). LFI-CAM: Learning Feature Importance for Better Visual Explanation.
In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1355-1363).

LFI-CAM

LFI-CAM 구조

- ❖ Perception Branch와 Attention Branch로 나누어진 모델 구조를 지님
- ❖ Perception Branch: 요약된 특징 맵 추출 및 분류 예측 수행
- ❖ Attention Branch: Attention map을 추출해 분류 모델 성능 향상에 도움을 줌

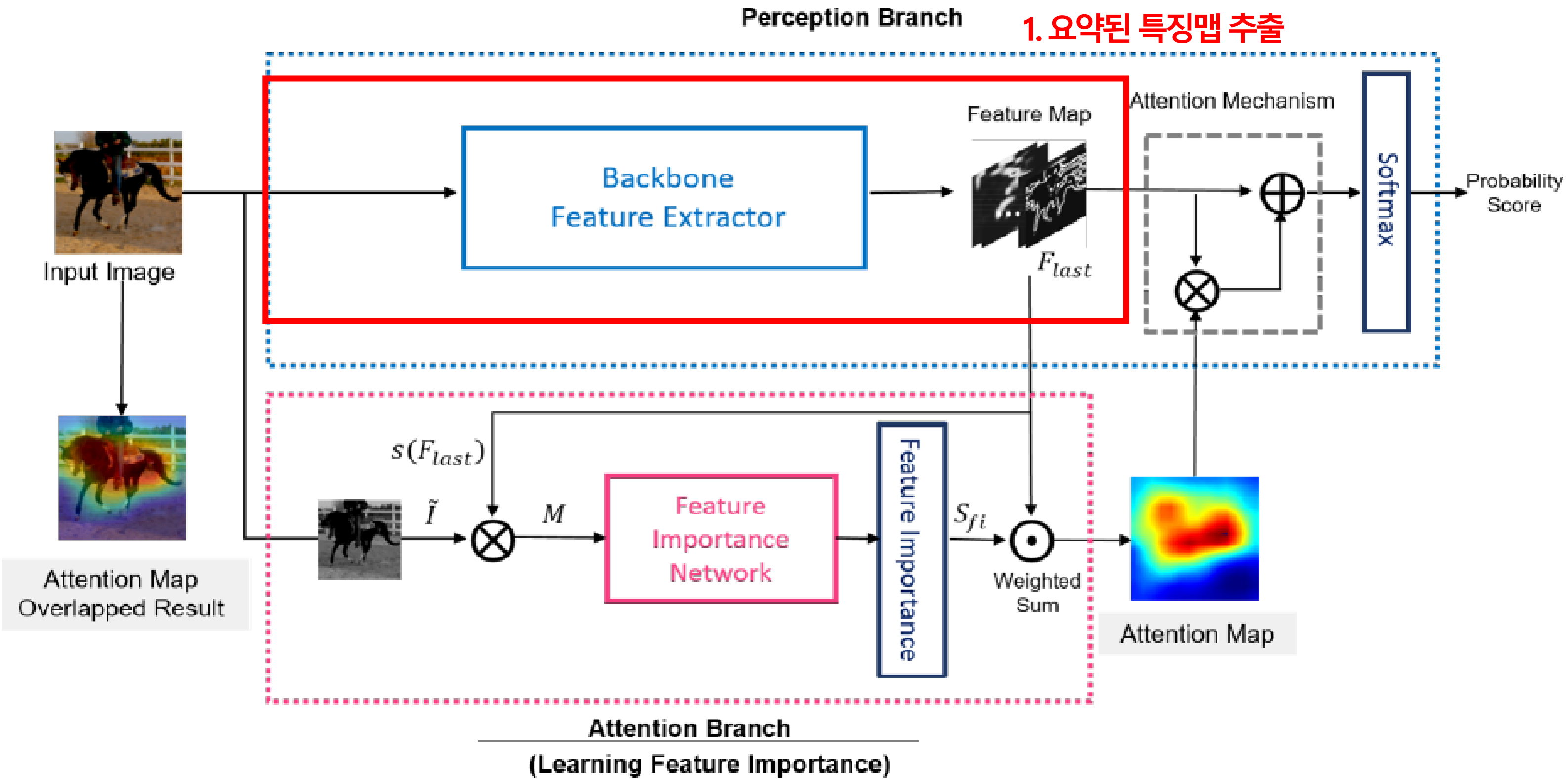


LFI-CAM

LFI-CAM 구조

❖ Perception Branch: 요약된 특징 맵 추출 및 분류 예측 수행

1. 일반적인 CNN 모델에서 마지막 Convolutional Layer를 통과한 특징맵을 추출

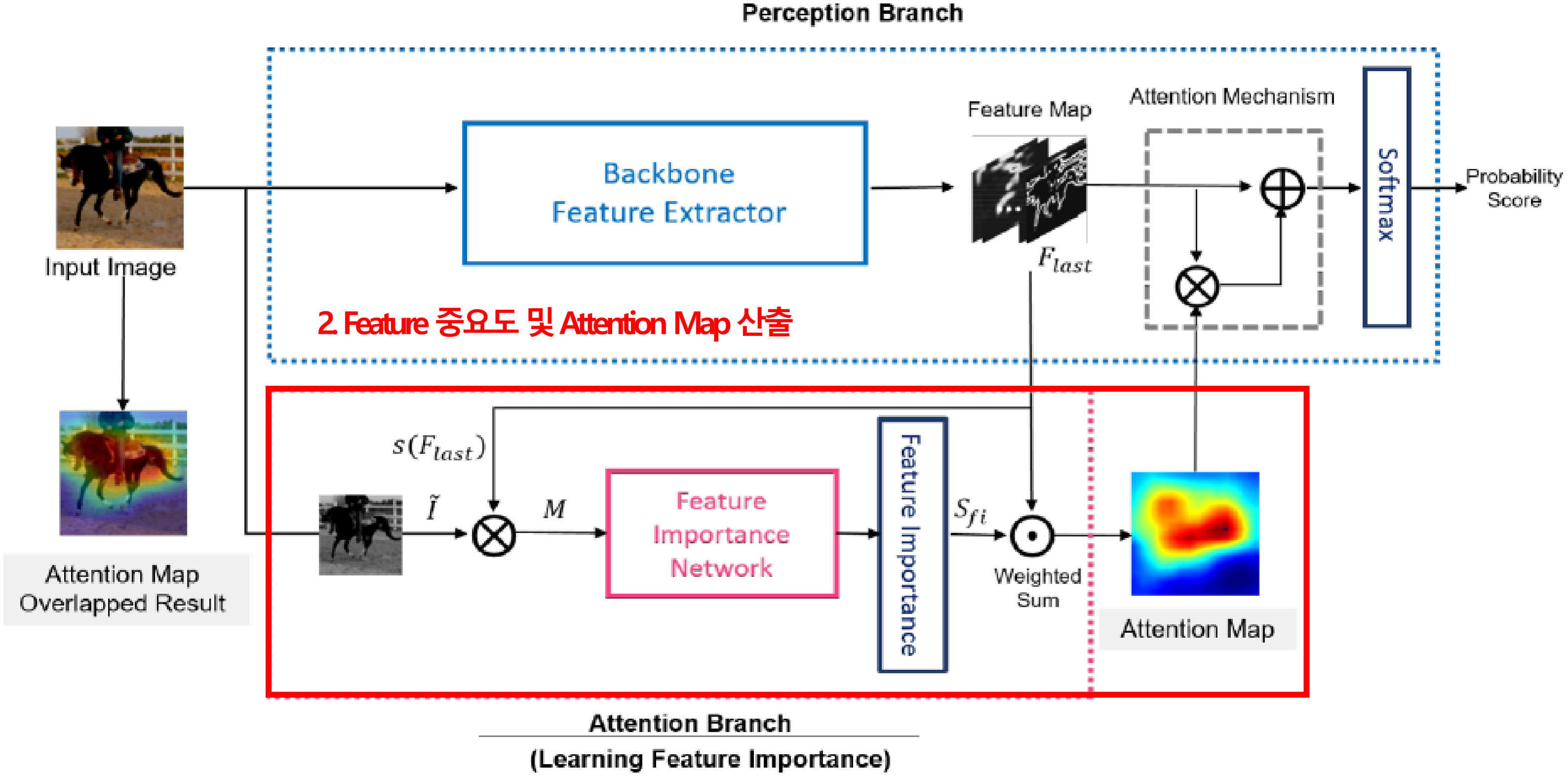


LFI-CAM

LFI-CAM 구조

❖ Attention Branch: Attention map을 추출해 분류 모델 성능 향상에 도움을 줌

2. 요약된 특징맵과 grayscale로 변형한 원본 이미지를 활용해 feature 중요도와 Attention map을 산출함

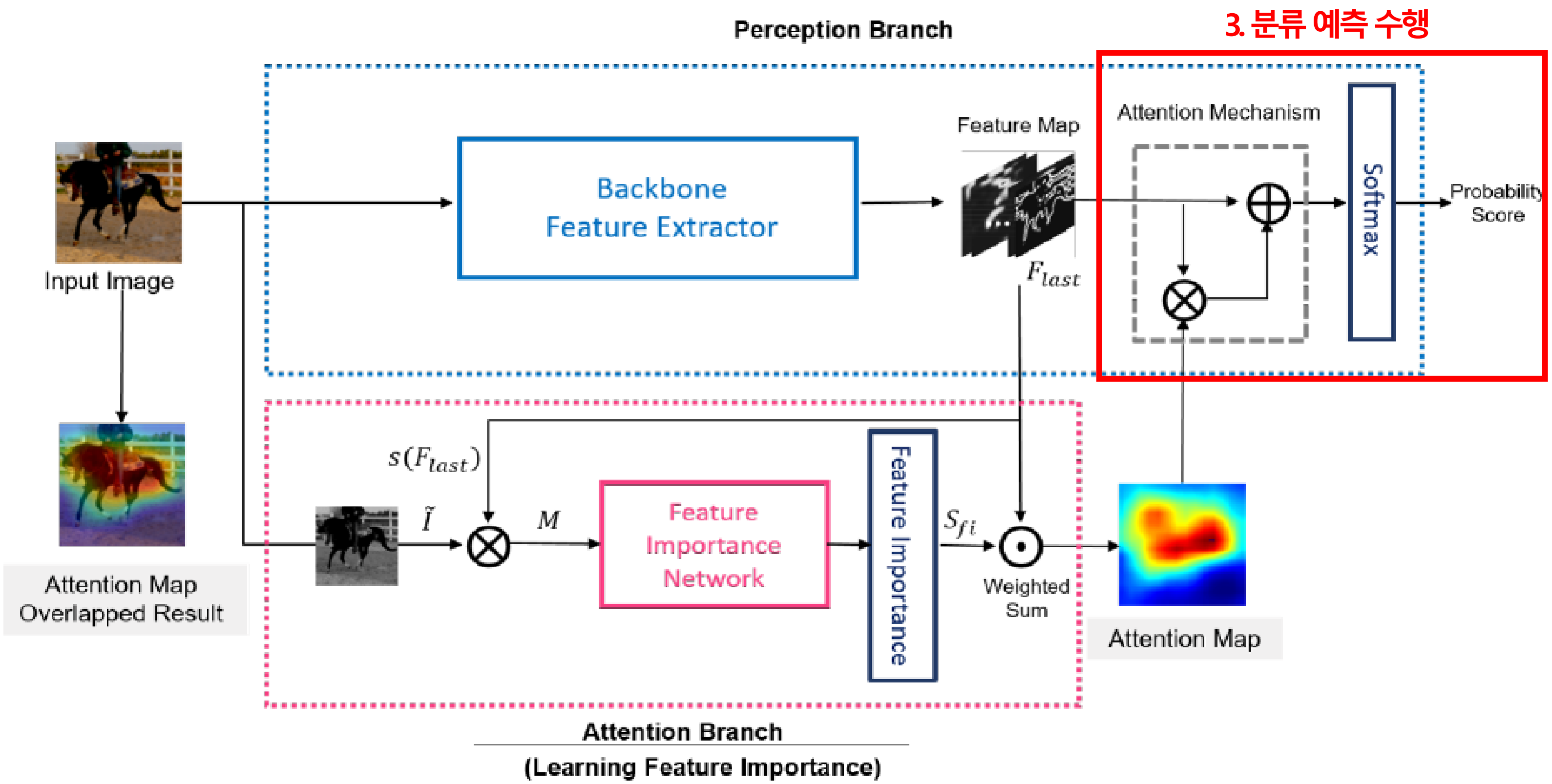


LFI-CAM

LFI-CAM 구조

❖ Perception Branch와 Attention Branch로 나누어진 모델 구조를 지님

3. 산출한 Attention map과 마지막 convolutional layer를 통과한 feature map을 활용해 분류 예측 수행



LFI-CAM

LFI-CAM 결과

❖ LFI-CAM 결과 – 단일 객체

- 기존 다양한 CAM 방식 중에서 예측에 중요했던 부분을 가장 잘 산출함을 확인

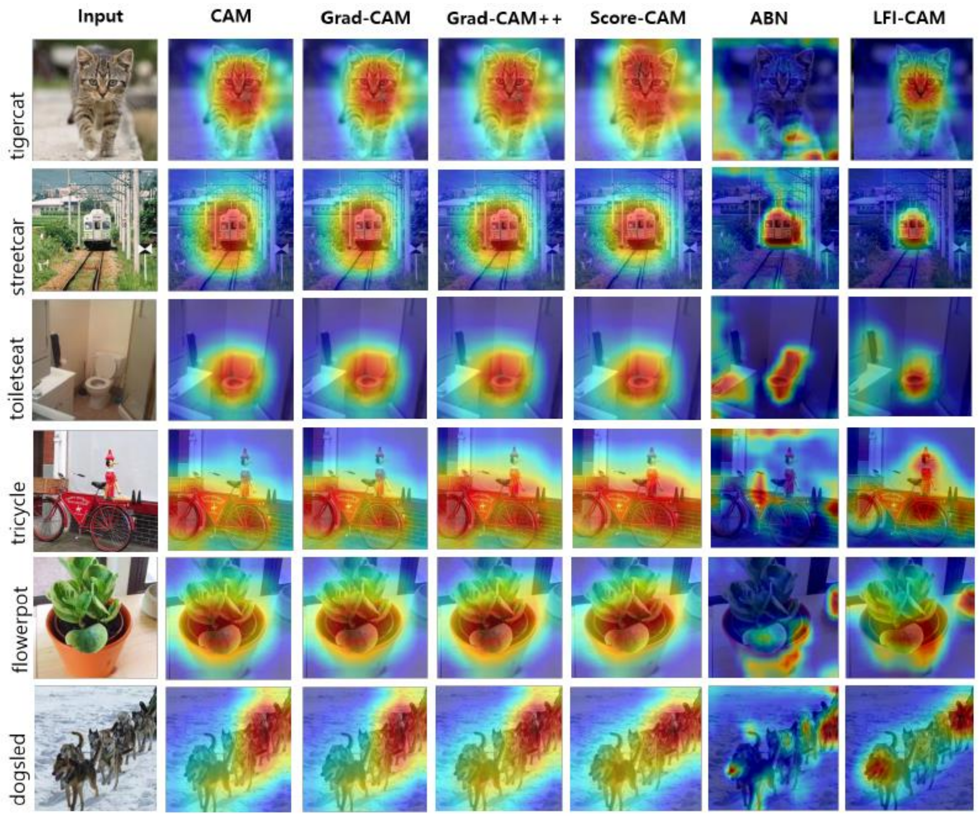


Figure 3. Visual explanation results of various methods on ImageNet. Notably, LFI-CAM always highlights the true class object correctly and in a more focused manner. For instance, LFI-CAM's tiger cat, streetcar, and toilet seat attention maps are tighter and focused on the salient features of the true class than any other methods. Additional results are provided in the supplementary material.

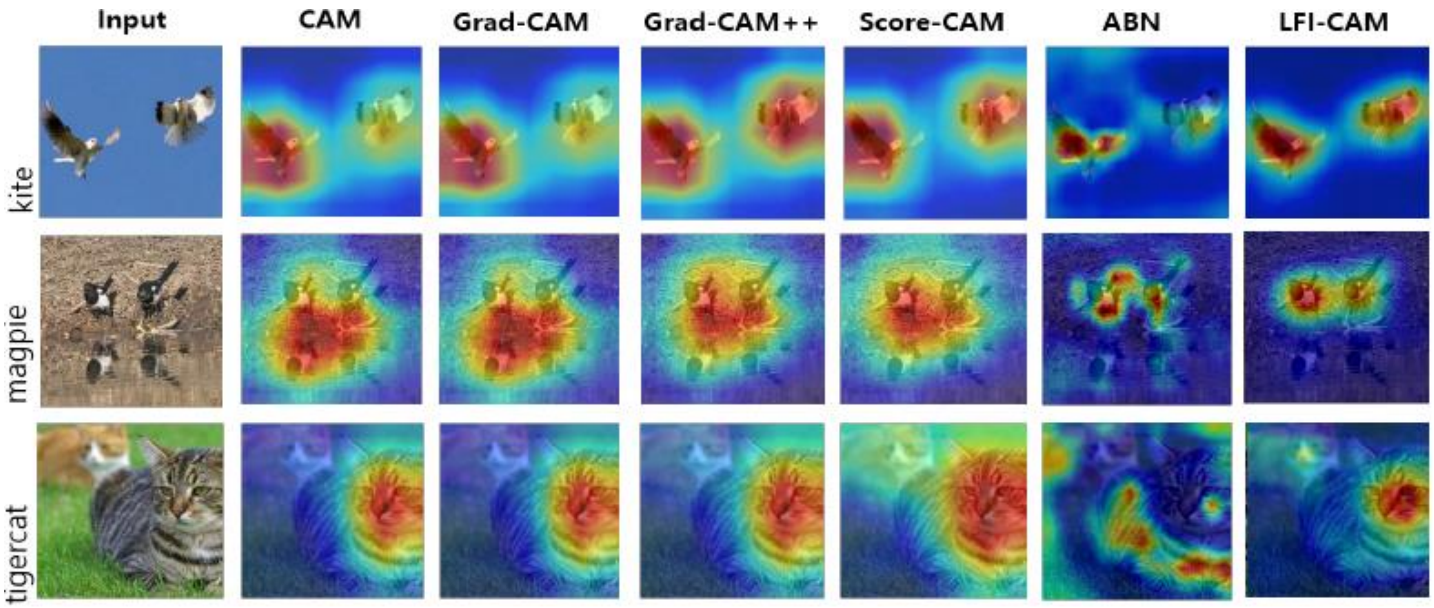
Reference: Lee, K. H., Park, C., Oh, J., & Kwak, N. (2021). LFI-CAM: Learning Feature Importance for Better Visual Explanation.
In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1355-1363).

LFI-CAM

LFI-CAM 결과

❖ LFI-CAM 결과 – 다중 객체

- 다중 객체에 대한 문제에서도 LFI-CAM이 다른 모델 대비 더 성능이 좋았음을 확인



Reference: Lee, K. H., Park, C., Oh, J., & Kwak, N. (2021). LFI-CAM: Learning Feature Importance for Better Visual Explanation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1355-1363).

7. Conclusions

Conclusions

❖ Conclusions

✓ 인공지능 모델이 발전함에 따라, 성능 뿐만 아니라 예측된 결과에 대한 근거를 해석할 수 있는 설명 가능한 AI (XAI)에 대한 중요도가 높아지고 있다.

✓ 모델에서 잘 요약된 feature map과 그 feature map의 중요도 (=weight)를 산출한다면 예측 결과에 대한 근거를 시각화해 보여줄 수 있다.

✓ 모델에 대한 해석을 위해 다양한 Class Activation Map (CAM) 알고리즘이 발전되어 왔다.

1) CAM: 중요도 산출을 위해 Global Average Pooling (GAP)을 활용해 연산함
2) Grad-CAM: GAP 구조 없이 딥러닝 모델의 gradient를 활용해 중요도를 산출함
3) Grad-CAM++: Gradient의 weighted average를 normalized term을 활용해 일반화된 Grad-CAM 제안
4) Score-CAM: Gradient 없이 채널별 중요도 차이 (CIC)를 활용해서 중요도를 산출함
5) LFI-CAM: Attention mechanism을 활용해 중요도를 산출함

✓ 알고리즘의 특징을 중심으로, image, signal, tabular data 등 여러 분야에 적용 가능할 것으로 기대

감사합니다.

Appendix

Data Mining
Quality Analytics

❖ Reference (Paper)

- Cheng, Chi-Tung, et al. "Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs." European radiology (2019): 1-9.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).
- Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE.
- Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., ... & Hu, X. (2020). Score-CAM: Score-weighted visual explanations for convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 24-25).
- Lee, K. H., Park, C., Oh, J., & Kwak, N. (2021). LFI-CAM: Learning Feature Importance for Better Visual Explanation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1355-1363).

❖ Reference (Website)

- https://www.datamaker.io/posts/17/
- https://alexisbcook.github.io/2017/global-average-pooling-layers-for-object-localization/
- https://jsideas.net/grad_cam/
- https://you359.github.io/cnn%20visualization/GradCAM/
- https://hugrypiggykim.com/2018/03/28/grad-cam-gradient-weighted-class-activation-mapping/
- https://velog.io/@tobigs_xai/CAM-Grad-CAM-Grad-CAMpp
- https://wewinserv.tistory.com/151
- https://jeonghwarr.github.io/review/score_cam/