



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

이상용 / 2020-03-16



Computational Data Science LAB



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

Computational Data Science LAB

목차

1. Introduction
2. Grad-CAM++ Methodology
3. Experiments



논의사항 및
결정사항

관련문서

Chattopadhyay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March).
In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 839–847). IEEE.



CONTENTS

1. Introduction
 2. Grad-CAM++ Methodology
 3. Experiments
- 
- 

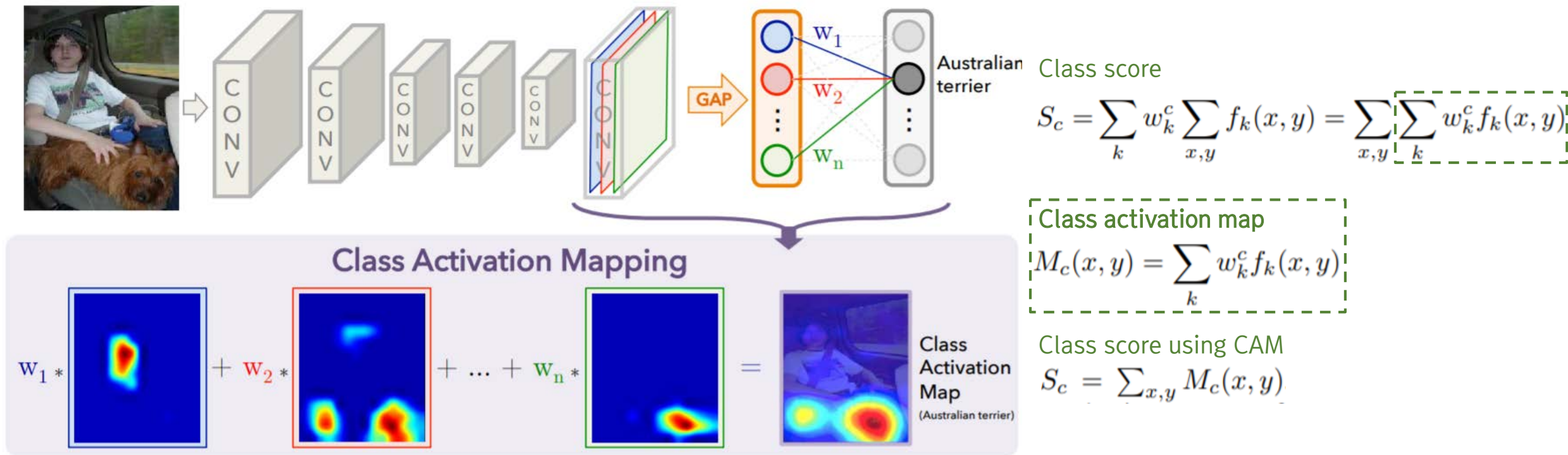
01 | Introduction

- 최근 10년 동안 CNN은 많이 복잡한 vision based problem들을 성공적으로 해결해옴
- 본 논문은 CNN 모델의 예측에 대한 더 나은 해석을 제공하기 위해 Grad-CAM++을 제안함
- Grad-CAM++은 Grad-CAM을 일반화한 방법이며 *localization of objects*와 *occurrences of multiple objects* 관점에서 Grad-CAM보다 더 나은 해석을 제공

01 | Introduction

CAM

- 마지막 Conv layer의 feature map에 GAP 적용
- Feature vector와 weight의 곱을 통하여 class activation map을 볼 수 있음



01 | Introduction

Grad-CAM

- 기존 CAM의 GAP를 해야만 하는 구조적 제약 없음 → 어느 Conv layer에서도 activation map 확인 가능
- CAM은 학습을 통해 얻은 weight를 사용 → backpropagation을 통한 gradient를 weight로 사용

Neuron importance weights

$$\alpha_k^c = \overbrace{\frac{1}{Z} \sum_i \sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$

Grad-CAM

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\underbrace{\sum_k \alpha_k^c A^k}_{\text{linear combination}} \right)$$

02 | Grad-CAM++ Methodology

- 앞선 방법들에서 특정 클래스에 대한 saliency map을 구하기 위해 다음과 같은 식을 사용

Saliency map

$$L_{ij}^c = \sum_k [w_k^c] A_{ij}^k$$

A_{ij}^k : k 번째 채널의 feature map

w_k^c : k 번째 채널에 대한 가중치

- Grad-CAM++ 또한 다른 방식으로 weight를 구함

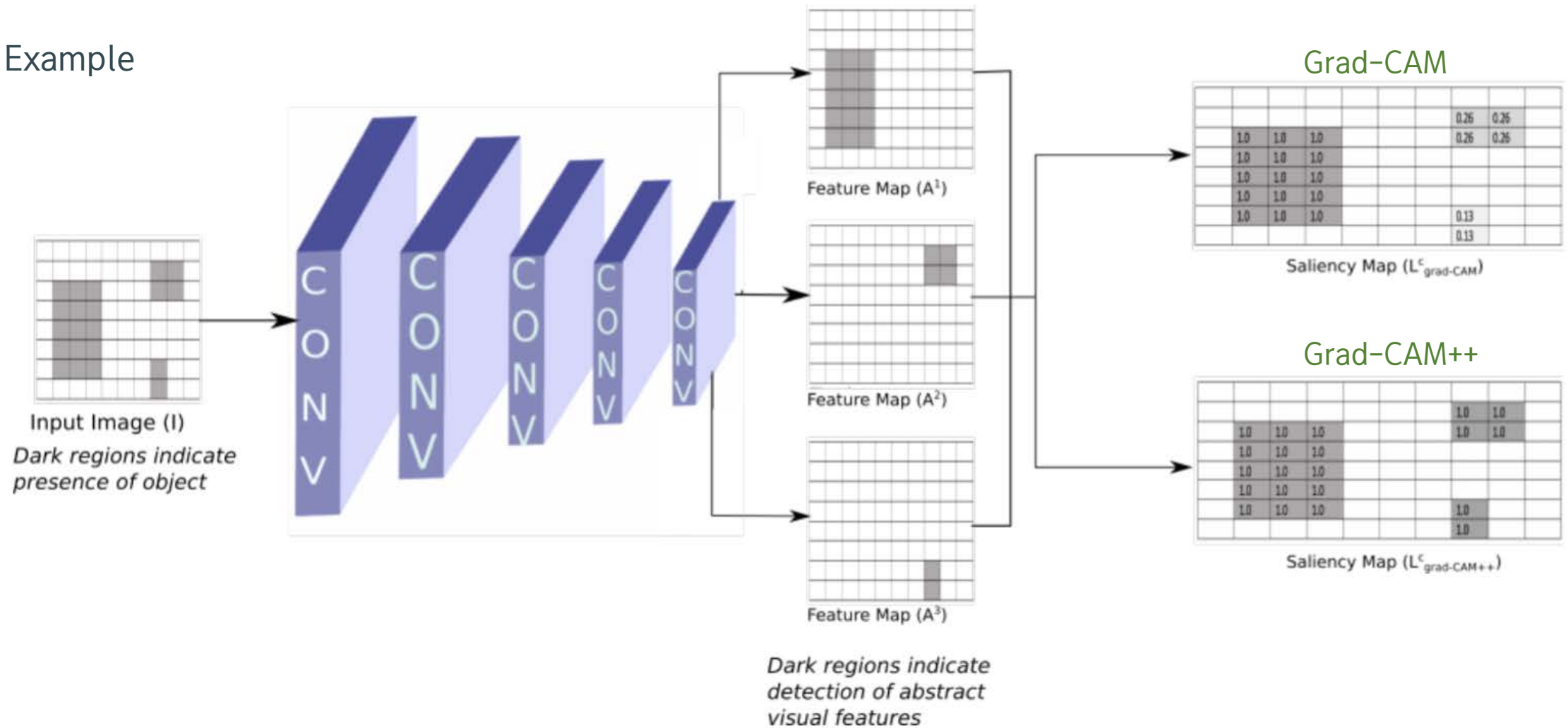
Grad-CAM++

$$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc} \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ij}^k}\right)$$

02 | Grad-CAM++ Methodology

Intuition

- Example



02 | Grad-CAM++ Methodology

Intuition

- 어떤 이미지 I 가 존재할 때, 그 이미지를 클래스 c 로 예측하게끔 영향을 준 픽셀에 대해 $A_{ij}^k = 1$ 이라고 가정, 그에 대한 미분값도 1이라고 가정

$$\begin{aligned}\frac{\partial y^c}{\partial A_{ij}^k} &= 1 && \text{if } A_{ij}^k = 1 \\ &= 0 && \text{if } A_{ij}^k = 0\end{aligned}$$

- 위 예제에 대한 Grad-CAM에서의 weights w 는 $w_1^c = \frac{15}{80}$, $w_2^c = \frac{4}{80}$, $w_3^c = \frac{2}{80}$ 처럼 구해짐
- 만약 위 예제의 이미지가 하나의 object에 대한 multiple occurrences를 가진 이미지라면, Grad-CAM의 경우 weight가 약한 부분은 final saliency map에서 사라질 수 있음
- 명백히 한 object에 대해 영향을 미치는 픽셀을 찾았음에도 채널별로 가중치가 곱해지기 때문에 영향을 덜 미치게 됨

02 | Grad-CAM++ Methodology

Intuition

- 반면 Grad-CAM++의 방법을 사용하면, 명백히 한 object에 대해 영향을 미치는 픽셀에 대해 좀 더 명확한 saliency map을 구할 수 있음

$$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc} \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ij}^k}\right)$$
$$\alpha_{ij}^{kc} = \frac{1}{\sum_{l,m} \frac{\partial y^c}{\partial A_{lm}^k}} \quad \text{if } \frac{\partial y^c}{\partial A_{ij}^k} = 1$$
$$= 0 \quad \text{otherwise}$$

- Grad-CAM은 feature map에 대해서만 채널 별로 가중치를 다르게 준 반면, Grad-CAM++은 채널 별로 가중치(w_k^c)를 다르게 주는 동시에 미분값에 대해서도 가중치(α_{ij}^{kc})를 달리 줌
- 따라서, 예제에 대한 Grad-CAM++에서의 weights w 는 $w_1^c = w_2^c = w_3^c = 1$ 로 동일함

02 | Grad-CAM++ Methodology

- CAM, Grad-CAM에서 사용하는 class score의 식은 다음과 같음

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y)$$

- Grad-CAM++ 도 위의 식처럼 변환

$$Y^c = \sum_k \left[\underbrace{\left\{ \sum_a \sum_b \alpha_{ab}^{kc} \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ab}^k}\right) \right\}}_{\text{Grad-CAM++ weight}} \left[\sum_i \sum_j A_{ij}^k \right] \right] = \sum_k \left[\sum_i \sum_j \left\{ \sum_a \sum_b \alpha_{ab}^{kc} \cdot \text{relu}\left(\frac{\partial Y^c}{\partial A_{ab}^k}\right) \right\} A_{ij}^k \right]$$

- ✓ (a,b) 는 A^k 에 대한 (i,j) 와 동일한 iterator이며 혼동을 방지하기 위하여 사용
- ✓ Gradient $\frac{\partial Y^c}{\partial A_{ij}^k}$ 에 대한 가중치를 달리 주기위해 α_{ij}^{kc} 를 구하기만 하면 됨

02 | Grad-CAM++ Methodology

Methodology

- α_{ij}^{kc} 를 구하기 위해 A_{ij}^k 에 대하여 편미분 시행 (ReLU는 단순한 threshold이기 때문에 없음)

$$\frac{\partial Y^c}{\partial A_{ij}^k} = \sum_a \sum_b \alpha_{ab}^{kc} \cdot \frac{\partial Y^c}{\partial A_{ab}^k} + \sum_a \sum_b A_{ab}^k \left\{ \alpha_{ij}^{kc} \cdot \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} \right\} \longrightarrow \text{Feature map에 대한 변화량}$$

- A_{ij}^k 에 대하여 편미분 시행하여 이계도함수 구함

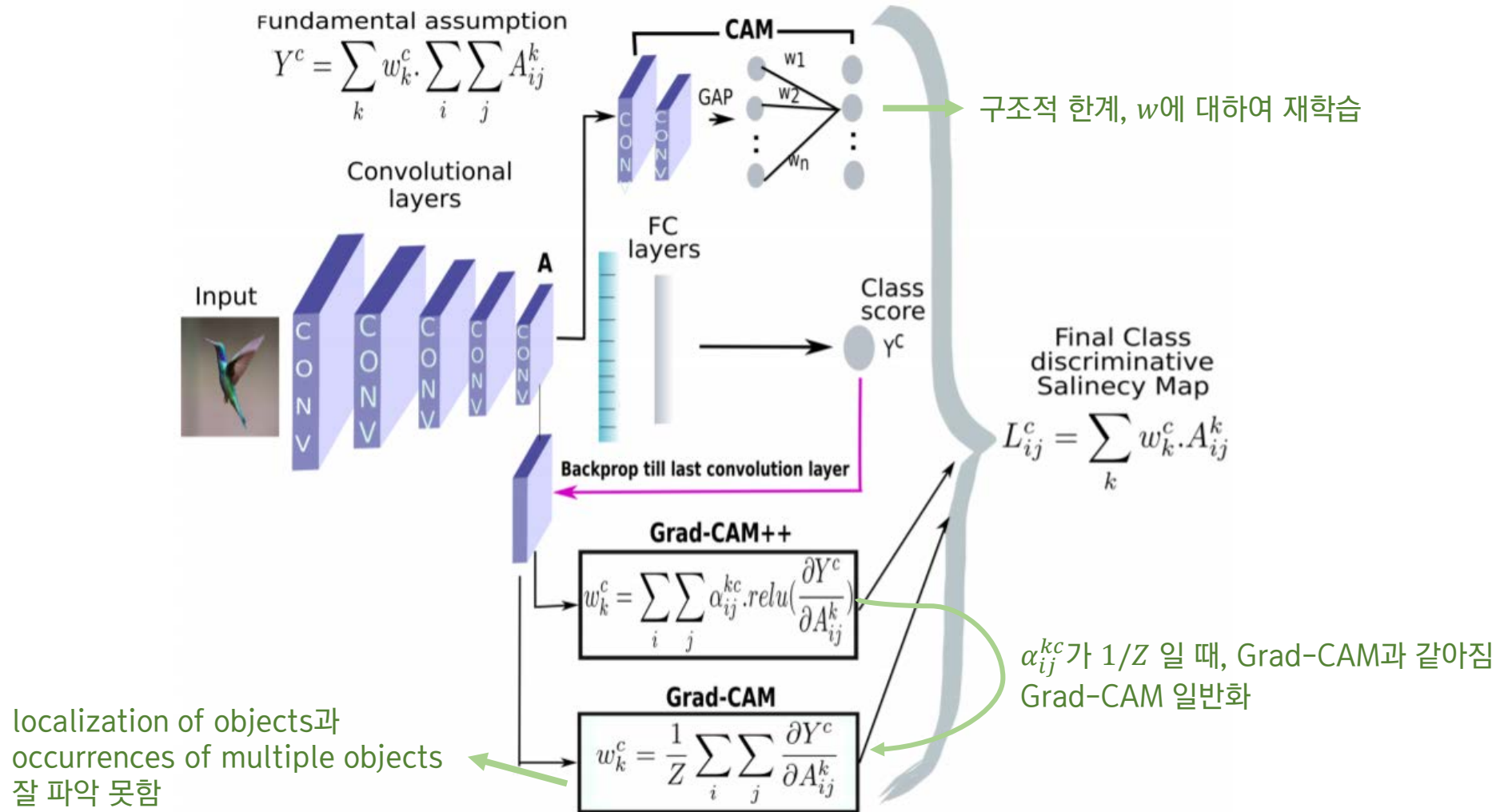
$$\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} = 2 \cdot \alpha_{ij}^{kc} \cdot \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \left\{ \alpha_{ij}^{kc} \cdot \frac{\partial^3 Y^c}{(\partial A_{ij}^k)^3} \right\} \longrightarrow \text{Gradient에 대한 변화량 (gradient의 가중치를 구하기 위해)}$$

- α_{ij}^{kc} 에 대하여 정리

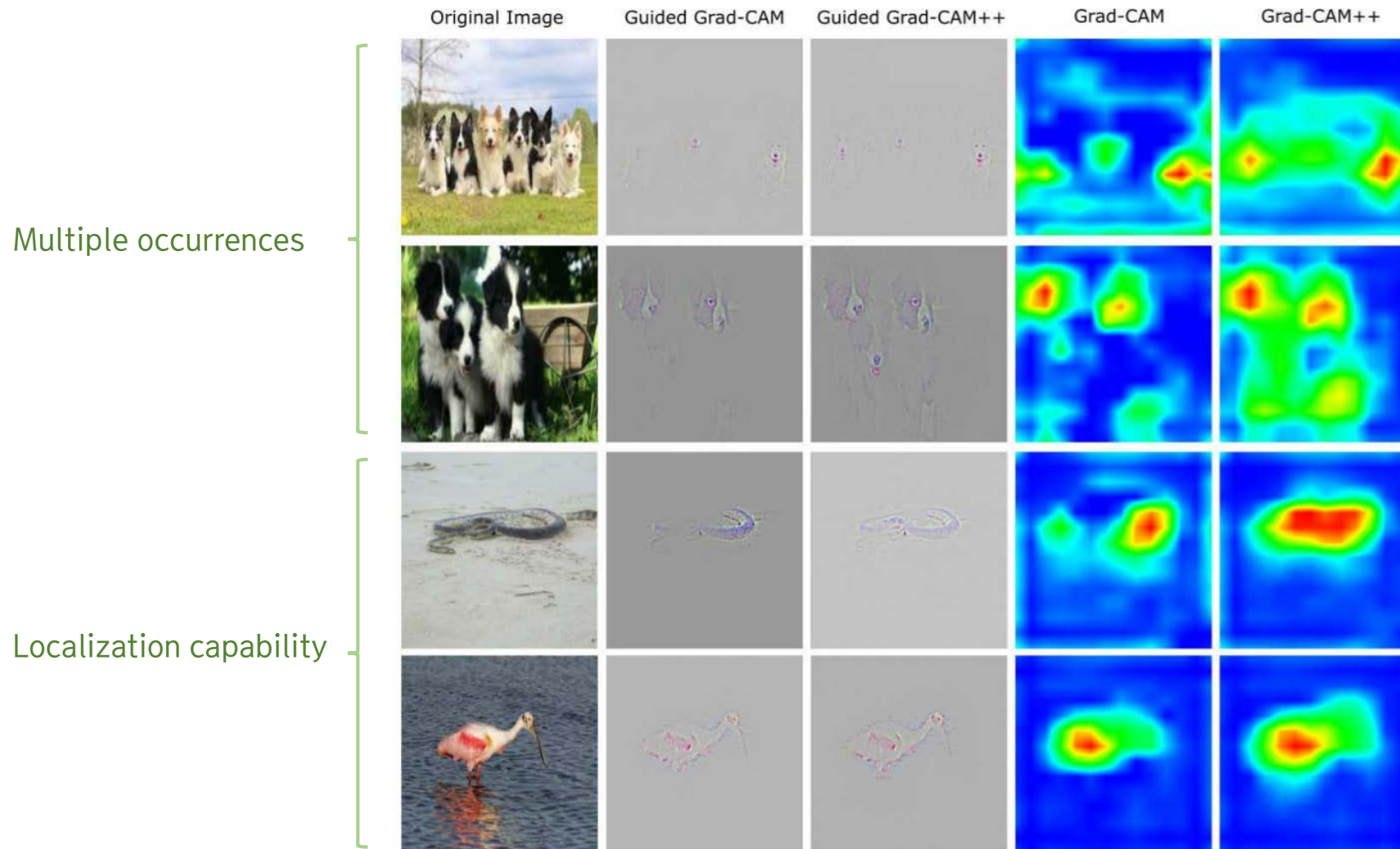
$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2}}{2 \frac{\partial^2 Y^c}{(\partial A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \left\{ \frac{\partial^3 Y^c}{(\partial A_{ij}^k)^3} \right\}}$$

02 | Grad-CAM++ Methodology

Overview of three methods

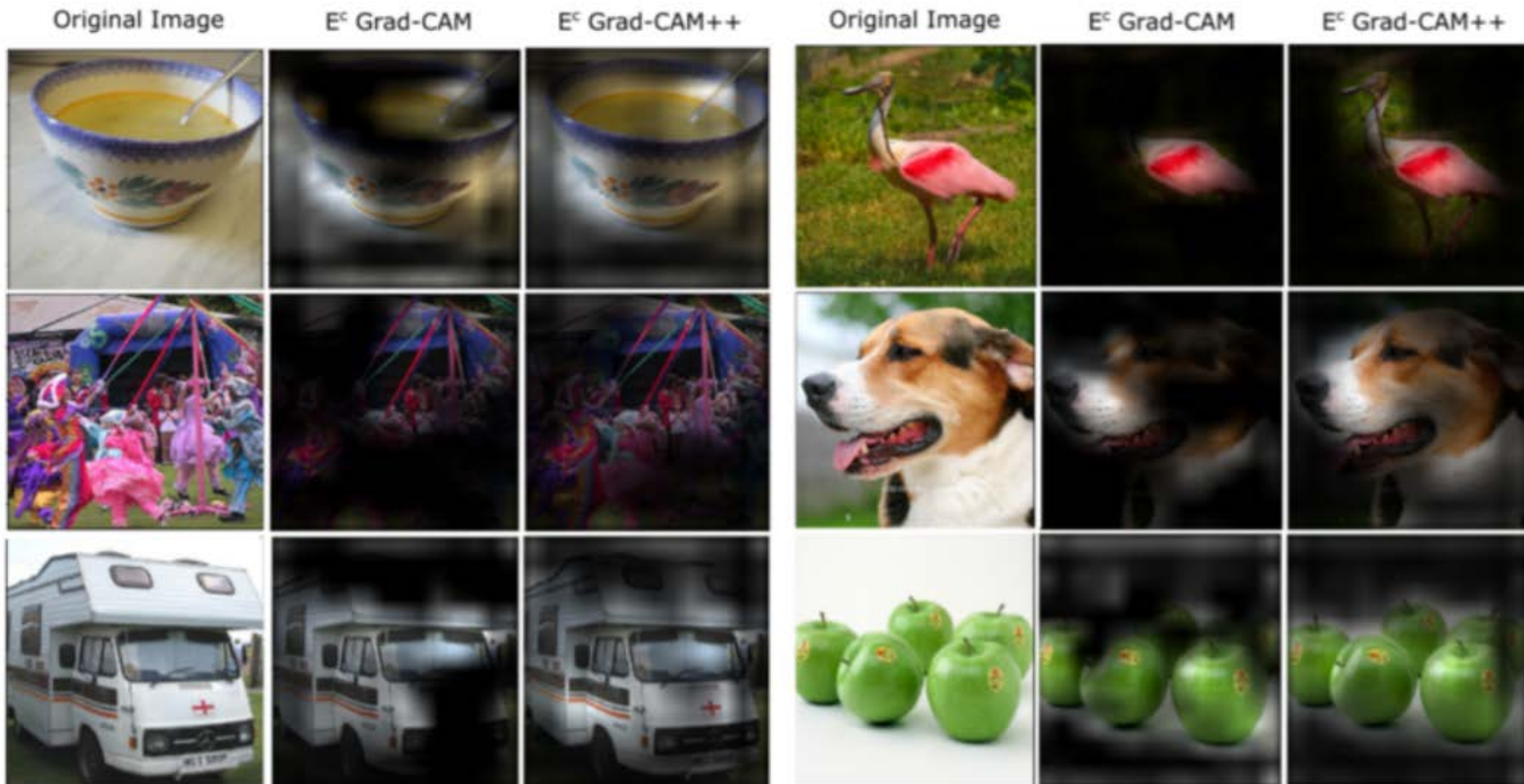


03 | Experiments



03 | Experiments

$$E^c = L^c \circ I$$



Q&A

감사합니다.