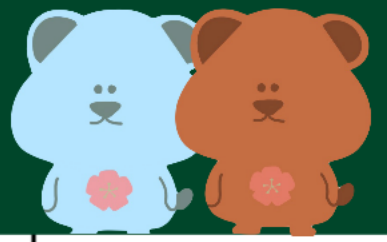


Relevance-CAM:
Your Model Already Knows Where to Look

Euron 6기 송윤진

EWHA EURON

목차

EWHA EURON

Introduction

EWHA EURON

01 Abstract

Relevance-weighted Class Activation Mapping(RelevanceCAM)

Layer-wise Relevance Propagation 활용

중간 레이어 분석 가능과 같은 장점

이미지 처리 모델의 각 레이어가 클래스별 특징을 추출

01 Introduction

모델 결정 해석하는 과정에서 기존 방법들은 한계가 존재

Relevance – CAM

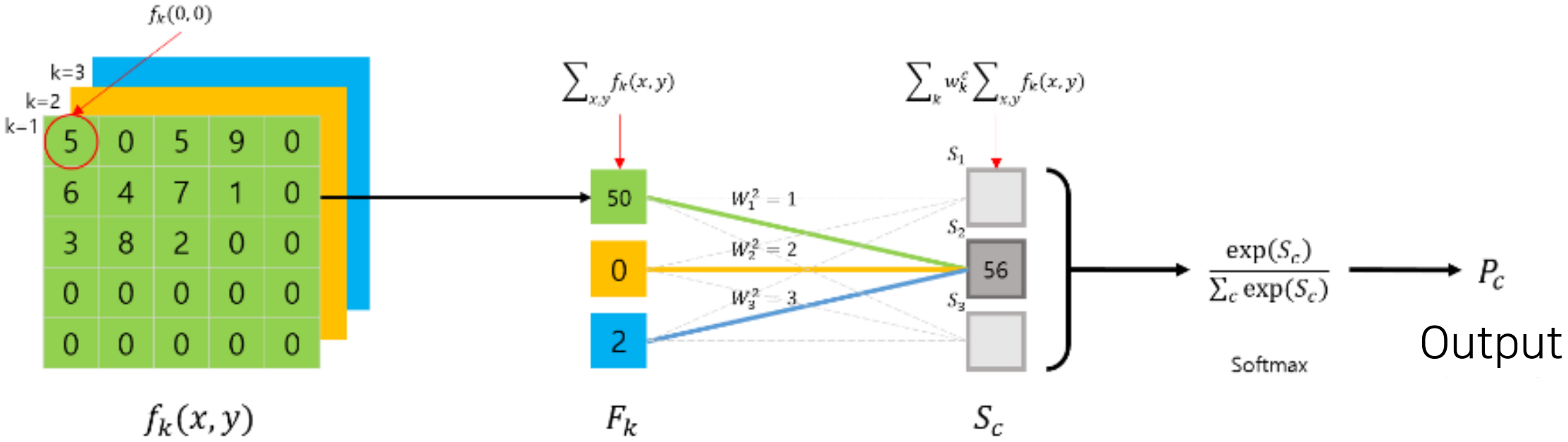
CAM은 히트맵을 통한 시각화 사용 → 더 정확하게 식별 가능

중간 레이어, 얕은 레이어에서 좋은 성능을 보임

Background

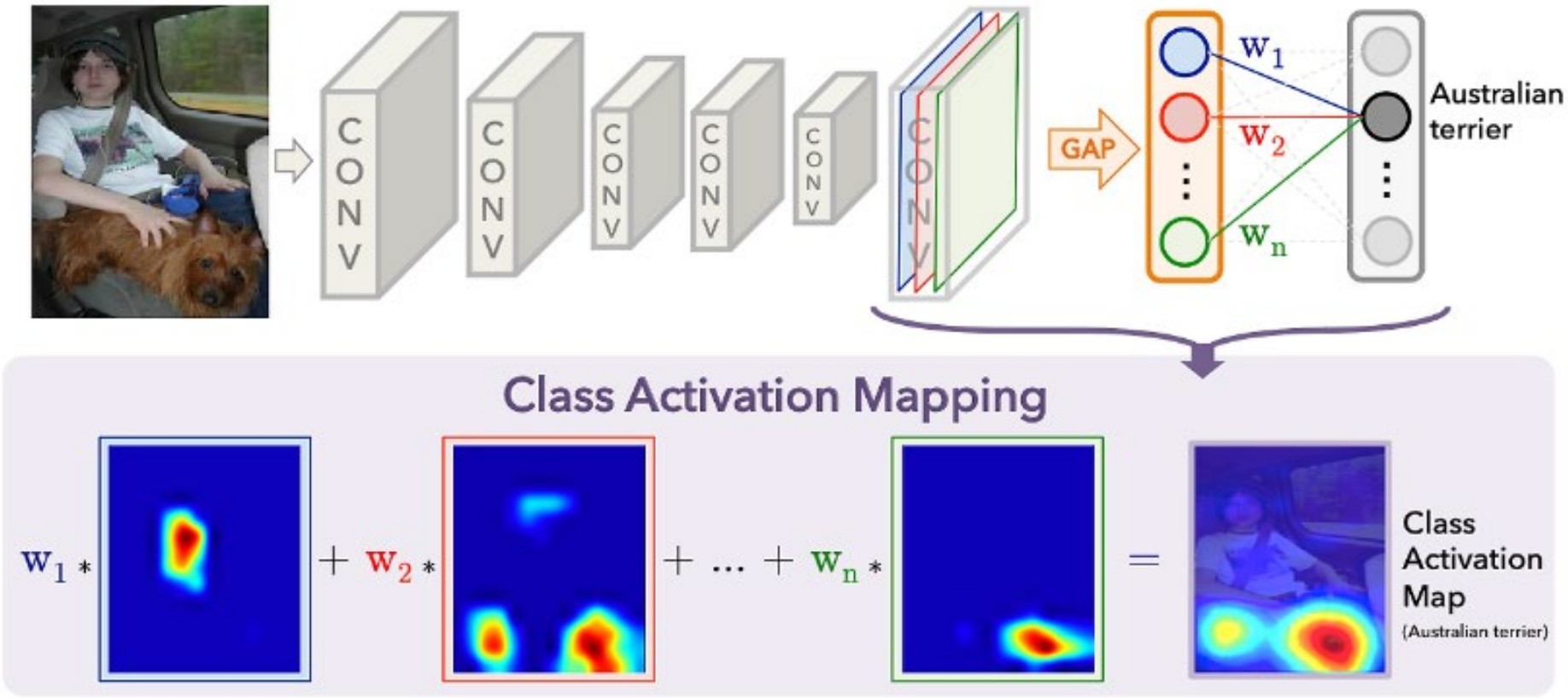
EWHA EURON

01 CAM


$f_k(0,0)$
k=3
k=2
k-1
5 0 5 9 0
6 4 7 1 0
3 8 2 0 0
0 0 0 0 0
0 0 0 0 0
$f_k(x,y)$
CNN Layer
$\sum_{x,y} f_k(x,y)$
50
0
2
F_k
GAP
$W_1^2 = 1$
$W_2^2 = 2$
$W_3^2 = 3$
$\sum_k w_k^c \sum_{x,y} f_k(x,y)$
S_1
S_2
56
S_3
S_c
$\frac{\exp(S_c)}{\sum_c \exp(S_c)}$
Softmax
P_c
Output

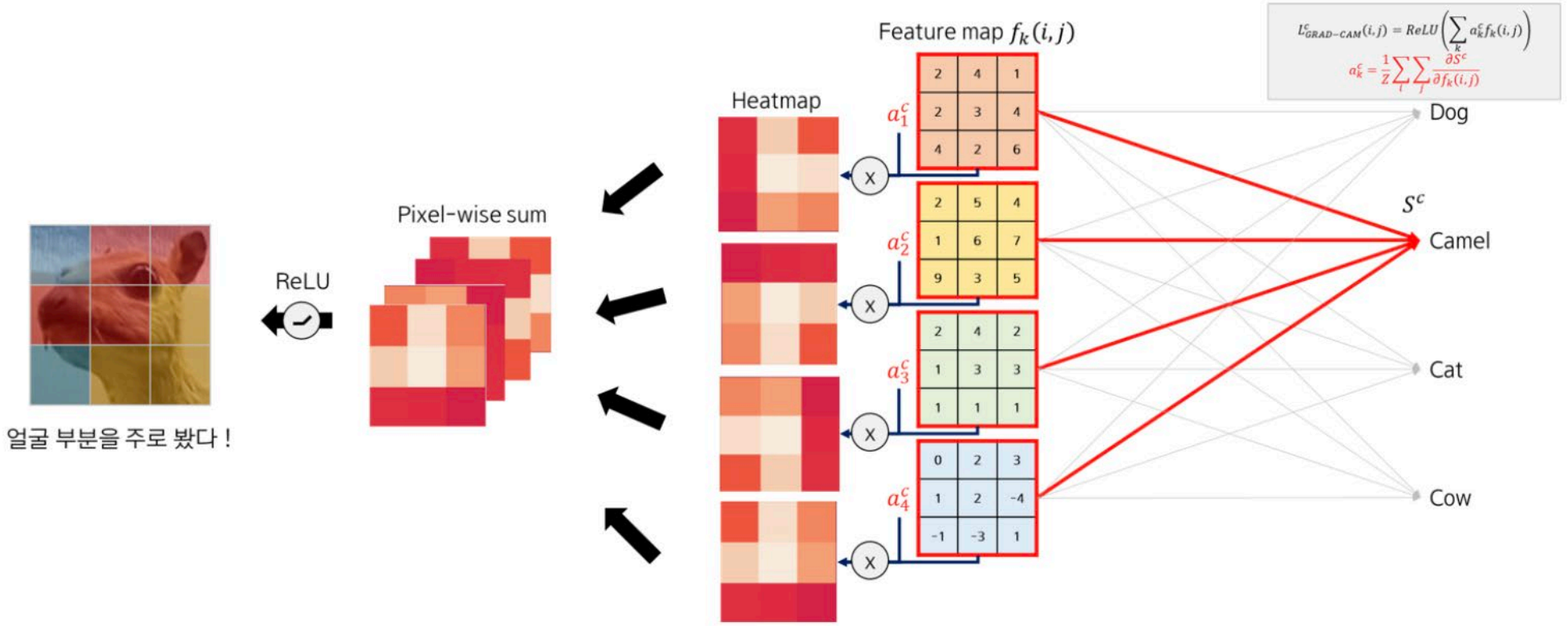
01 CAM

Class Activation Map: 어떻게 이미지를 특정 클래스로 예측했는지

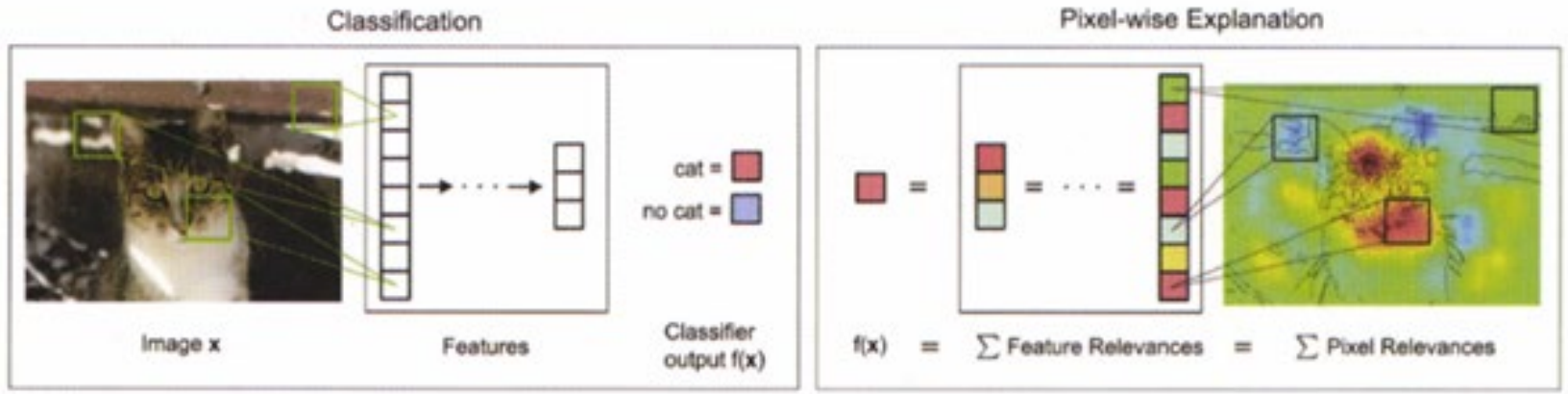


02 Grad-CAM

Gradient - CAM: 제한된 모델 구조, 분류 문제만 해석 가능한 한계를 해결



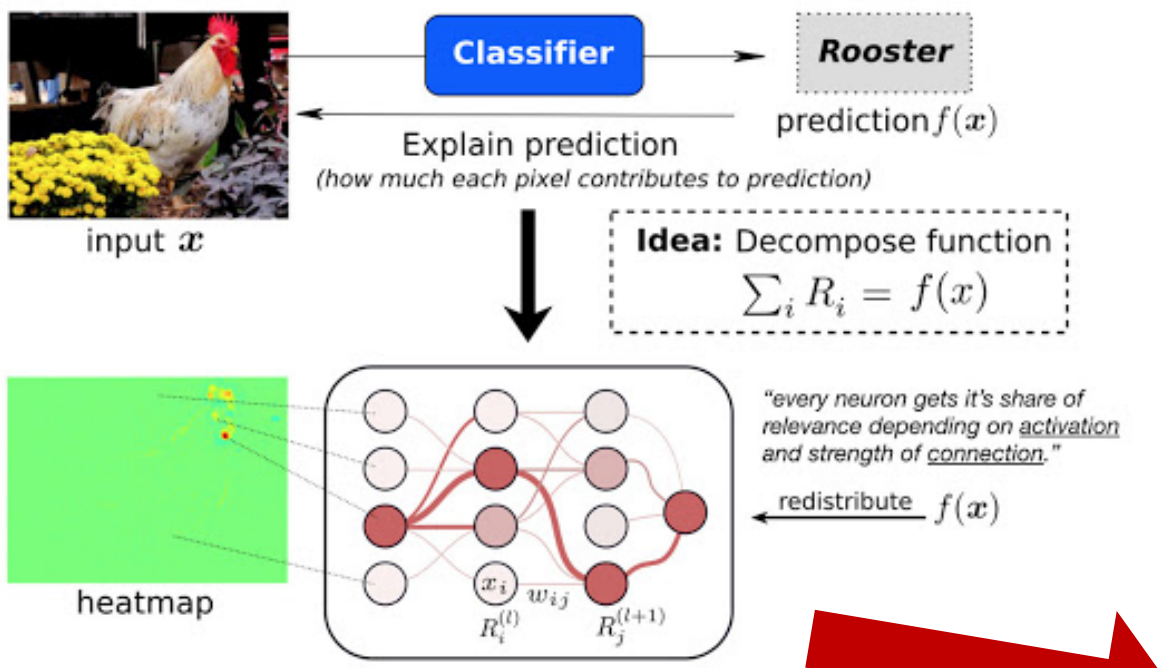
03 Layerwise Relevance Propagation(LRP)

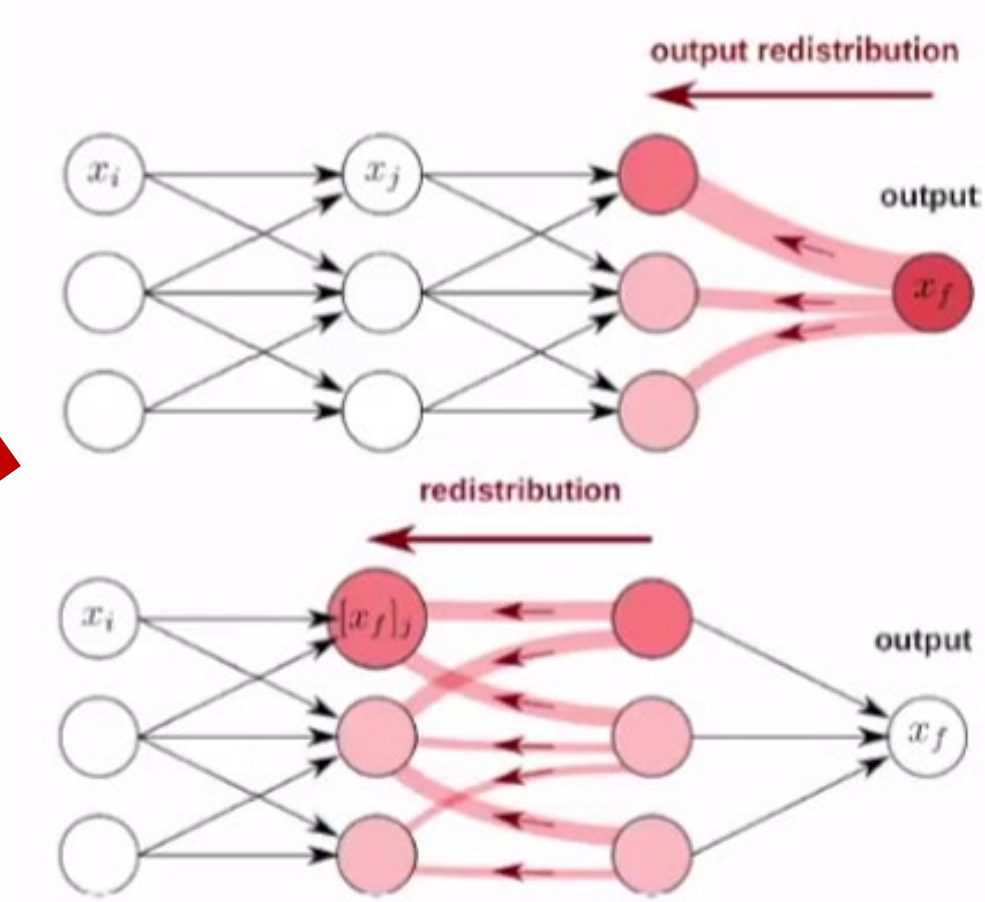


분해: 어떤 픽셀이 결과를 도출하는데 도움이 되는지(+) 안 되는지(-)

Relevance Propagation

: 결과값 출력에 어떤 기여를 하는지 관련성(Relevance)를 계산





03 Layerwise Relevance Propagation(LRP)

$$\forall x : f_c(x) = \sum_p R_p^l(x).$$

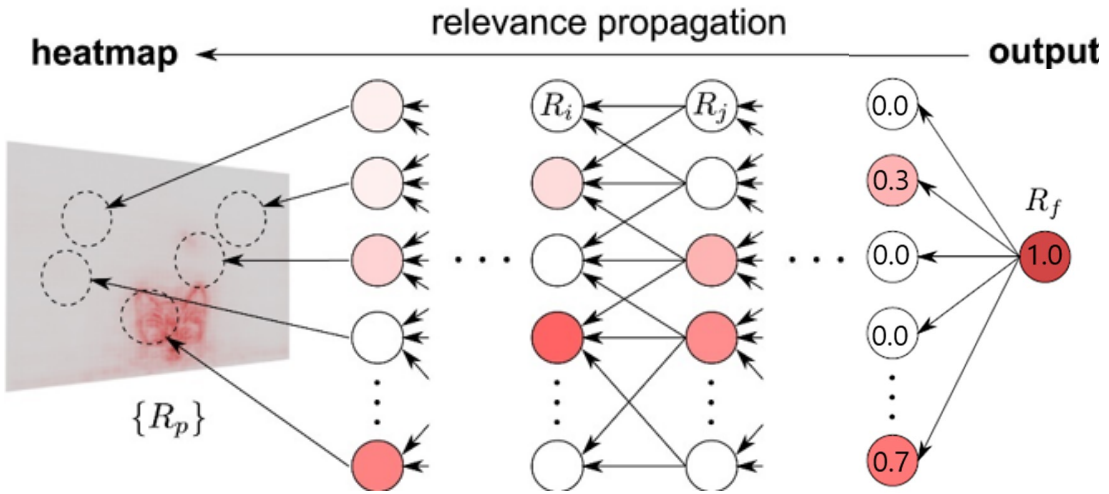
관련성 점수는 보존적 (ex. 재분배)

$$\forall x, p : R_p(x) \geq 0$$

관련성 점수가 양수인 경우

03 Layerwise Relevance Propagation(LRP)

$$R_i = \sum_j \frac{z_{ij}^+}{\sum_i z_{ij}^+} R_j$$



04 Contrastive Layerwise Relevance Propagation(CLRP)

$$R_n^{(L)} = \begin{cases} z_t^{(L)} & n = t \quad \text{타겟} \\ -\frac{z_t^{(L)}}{N-1} & otherwise \end{cases}$$

최종 레이어에서 타겟이 아닌 클래스의 관련성 감소

= 타겟 클래스에 대한 히트맵의 민감도 증가

05 Gradient Issue

1.Noisiness and discontinuity:

신경망이 깊어질 수록 노이지하고 불연속해짐

∵ pixel 별로 구해지기 때문에 인접한 pixel간의 관계성 확인하기 어려움

2. False Confidence:

Grad-CAM은 activation map의 출력 변화량(민감도)을 고려

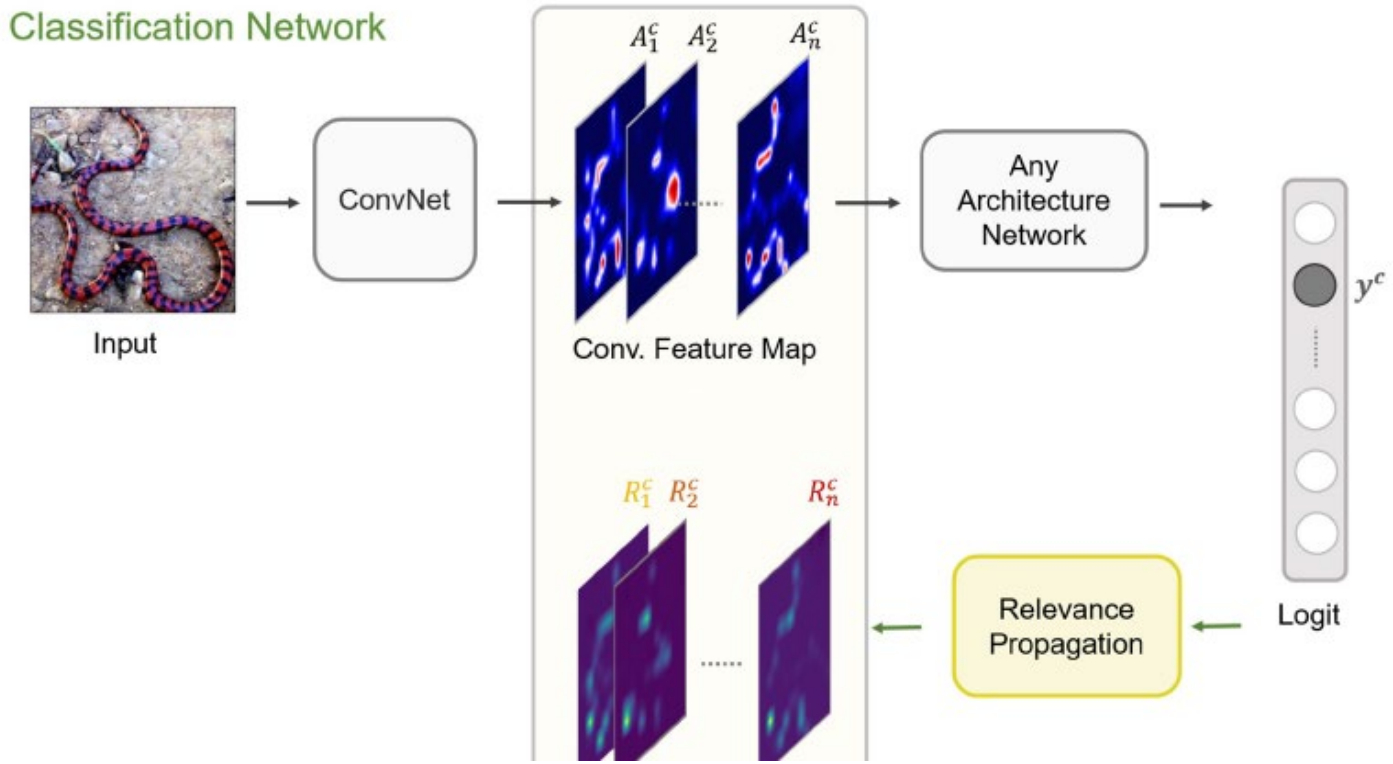
∴ 결과값에 대한 관련성(기여도)를 얻지 못함

Relevance Score: LRP의 관련성 점수를 CAM의 가중치 구성 요소로 고려

Relevance-CAM

Relevance-CAM

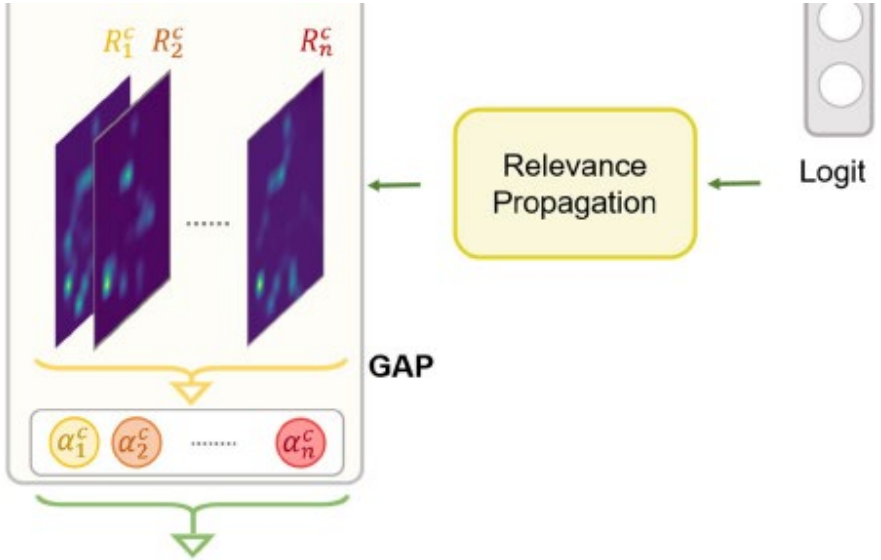
1. Forward Propagaion을 통한 activation map 구하기

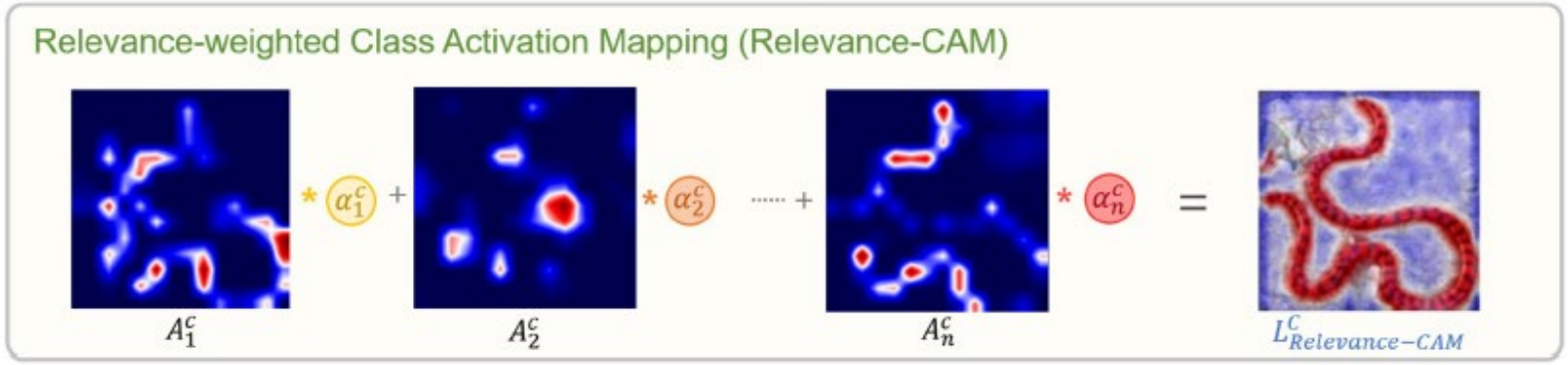


2. LPR을 따른 Backpropagation을 통한 관련성 점수 구하기

Relevance-CAM

3. 활성 가중치 구성 요소 GAP하기





4. Activation map과 가중치 곱하기

5. 히트맵 더하기

Relevance-CAM

i -th layer feature map k
for the target class

$$L^{(c,i)}_{Relevance-CAM} = \sum_k \alpha_k^{(c,i)} A_k^c$$

where

$$\alpha_k^{(c,i)} = \sum_{x,y} R_k^{(c,i)}(x,y)$$

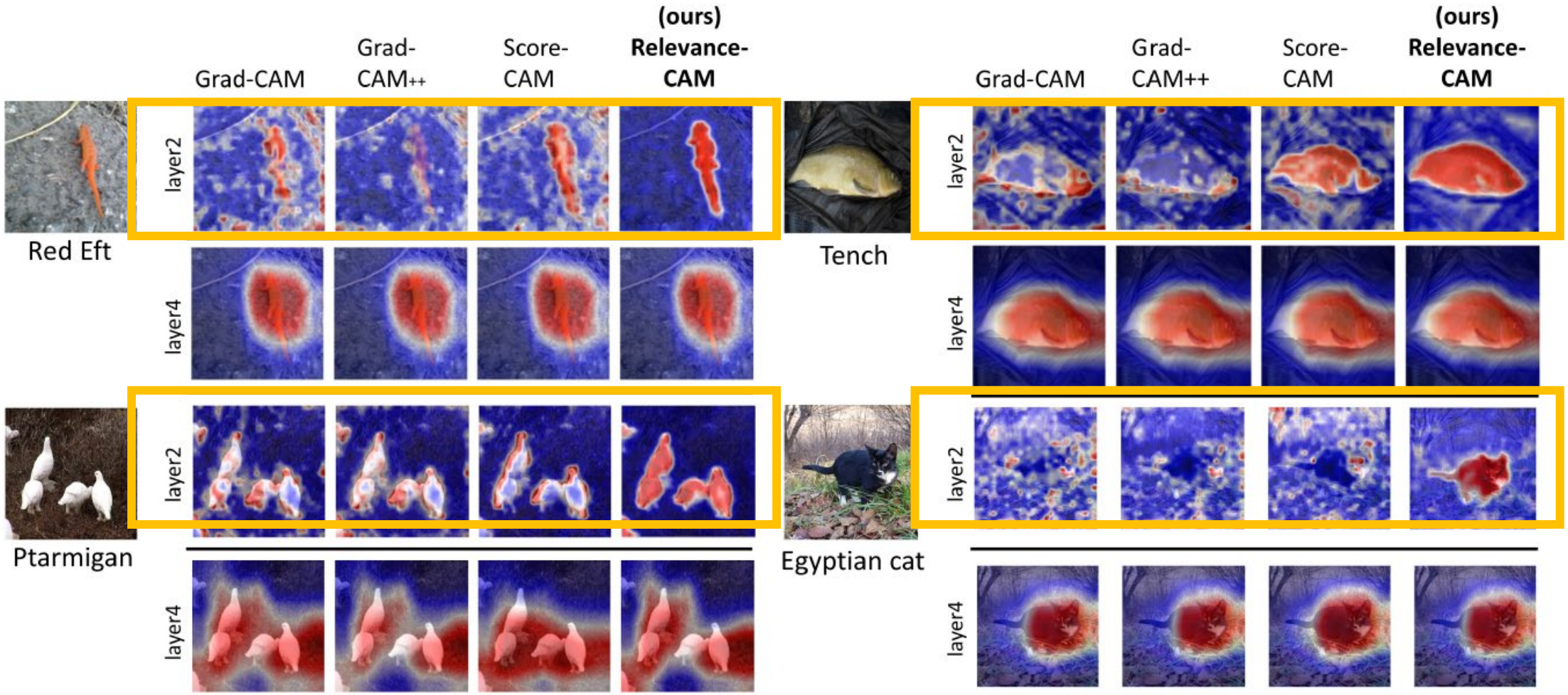
weighting component

relevance map of A_k^c

Experiment

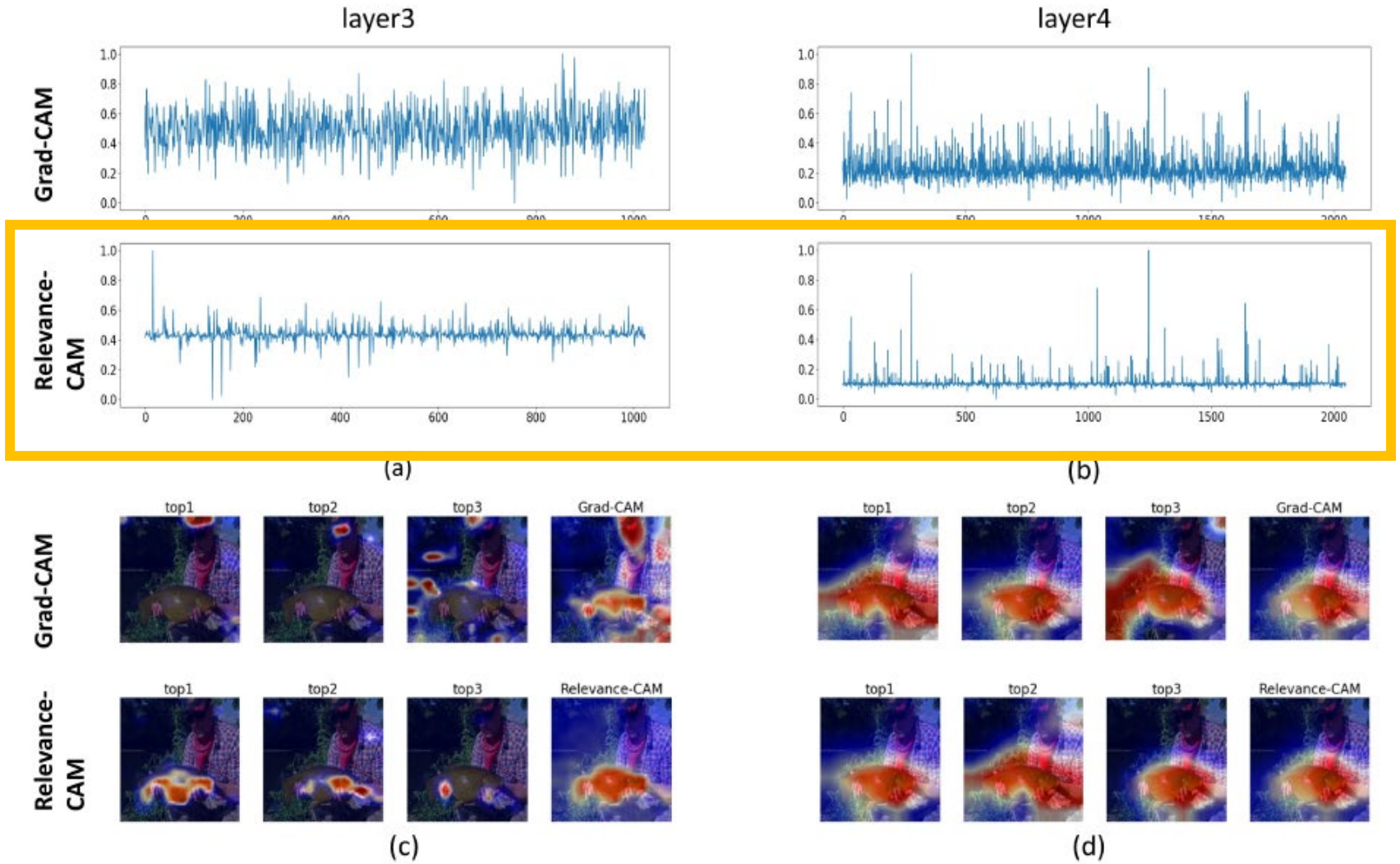
EWHA EURON

01 Depth-wise visualization

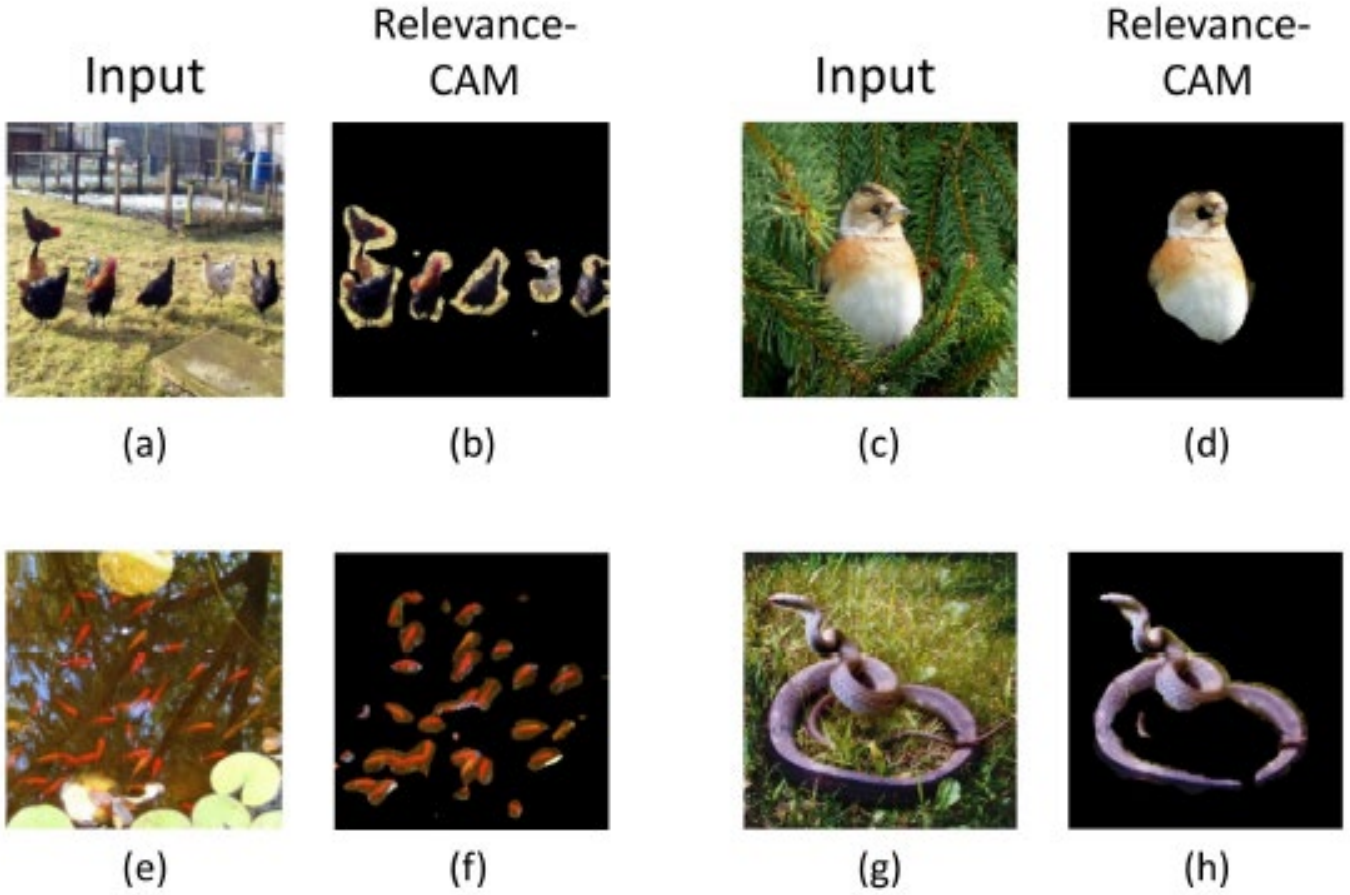


02 Evaluation for the selectivity

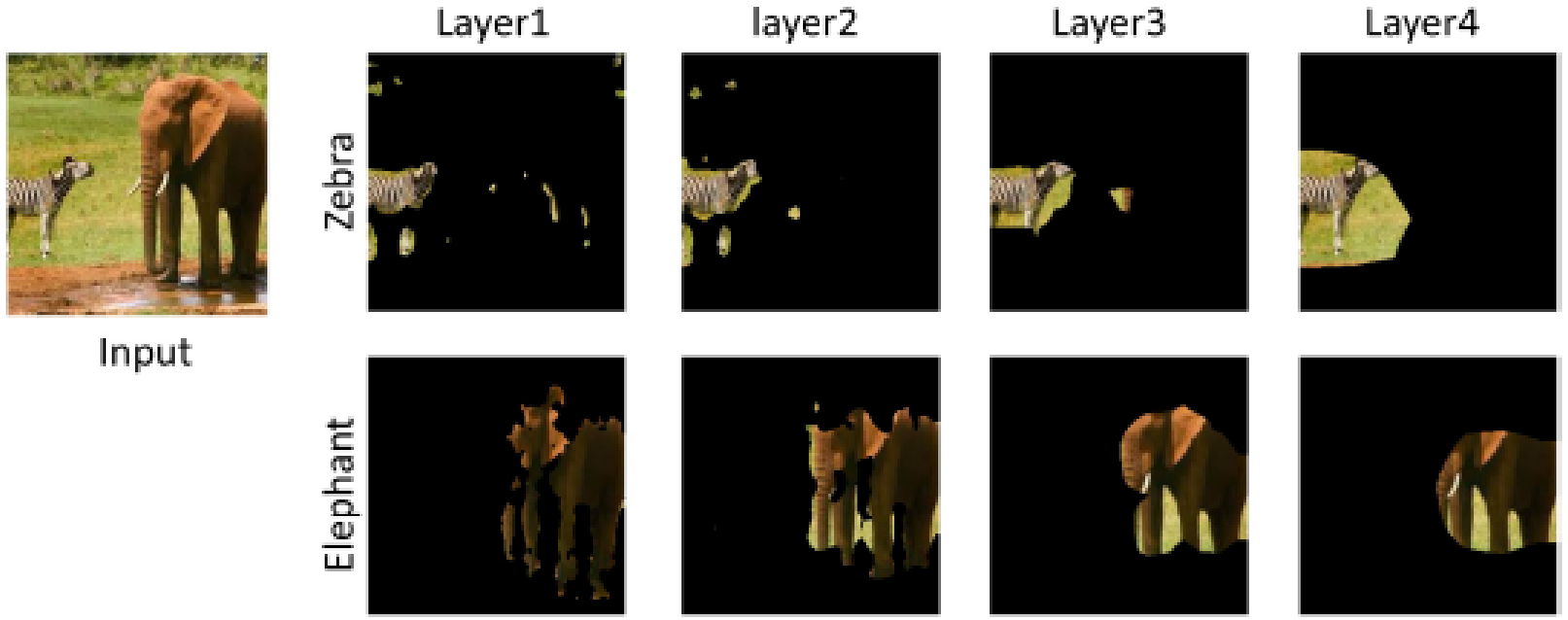
Input



03 Evaluation for Localization

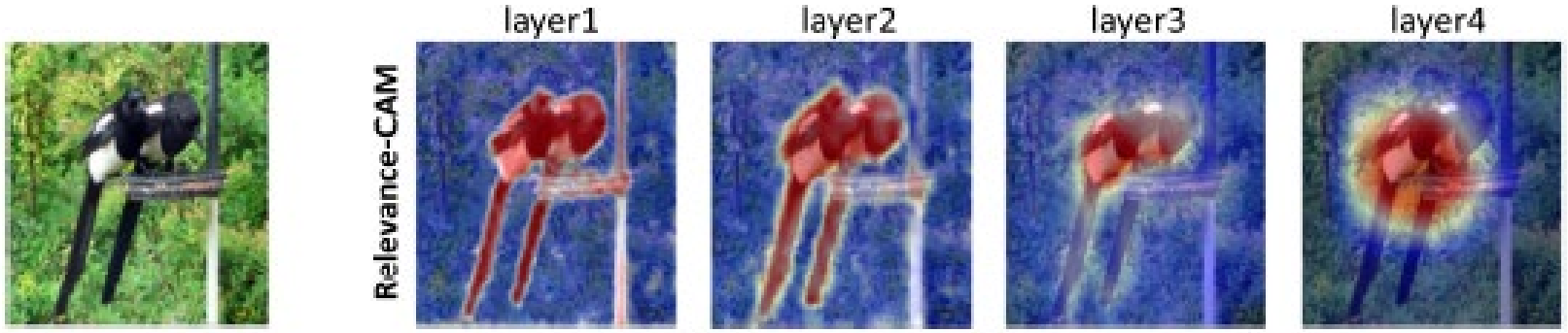


04 Class Sensitivity Test



	layer 1	layer 2	layer 3	layer 4
Grad-CAM	0.12	0.18	0.22	0.34
Grad-CAM++	0.13	0.19	0.22	0.34
Score-CAM	0.21	0.25	0.28	0.34
Relevance-CAM	**0.30**	**0.32**	**0.32**	0.34

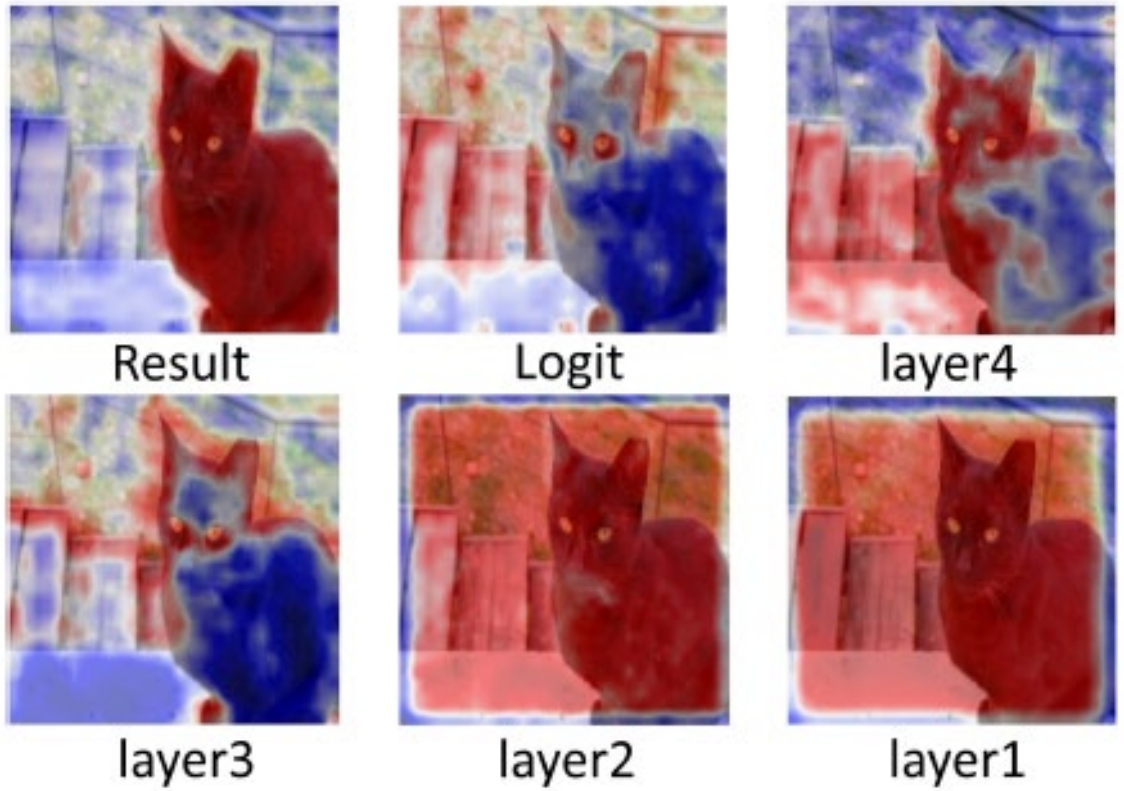
04 Class Sensitivity Test



Evidence that class specific information is extracted from shallow layers

: 얕은 레이어에서도 특정 클래스에 대한 정보를 추출할 수 있음

05 Sanity check for RelevanceCAM



Conclusion

EWHA EURON

Conclusion

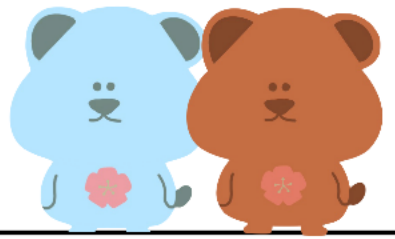
제안된 Relevance-CAM은

다양한 깊이의 레이어에서 신뢰성 있고 정확한 분석이 가능하다.

얕은 레이어에서도 클래스 특징 추출이 가능하다.

이를 이용하여 전이 학습, 약한 supervised learning에 적용이 가능하다

THANK YOU



EWHA EURON