

마스크 이미지를 이용한 CNN 기반 클래스 확률 보정

CNN-based Class Probability Correction using Mask-image

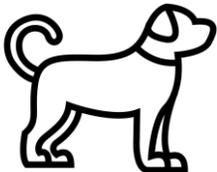
맹대주 석사과정, 백준걸 교수*

고려대학교 산업경영공학과

{eownaod, jungeol}@korea.ac.kr

1. Introduction

문제 정의 - Image Classification

Dog 



Cat 



Bird 



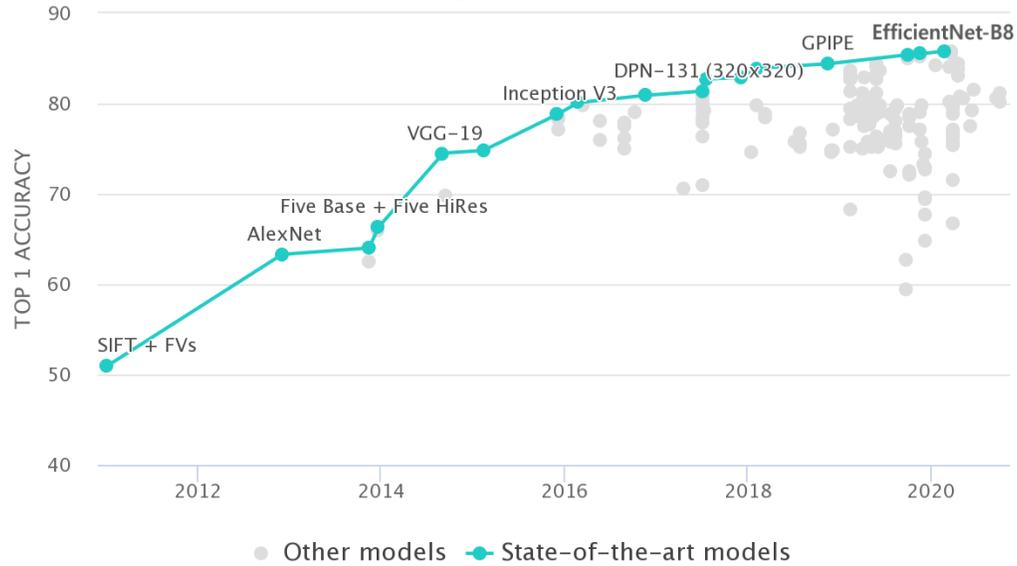
상위클래스

하위클래스

- 이미지 분류 문제에서 클래스는 상위 클래스 (Ex : 개, 고양이, 새 등)와 하위 클래스(Ex : 개의 품종 등)로 구분
- 상위 클래스는 클래스 간 특징이 명확
- 하위 클래스는 클래스 간 특징이 명확하지 않으며, 미세한 특징 차이만 존재
- 특히 수많은 상위 클래스와 하위 클래스가 혼재된 이미지 데이터는 분류 난이도가 높음

1. Introduction

문제 정의 - ImageNet 데이터



출처 : <https://paperswithcode.com/sota/image-classification-on-imagenet>

IMAGENET

Model	Top1-Acc.
EfficientNet-B8	85.8%
ResNeSt-269	84.5%
Assem-ResNet152	84.2%
ResNet152(Pytorch)	78.3%
VGG16(Pytorch)	71.6%



Kite



Bald Eagle



Fox terrier

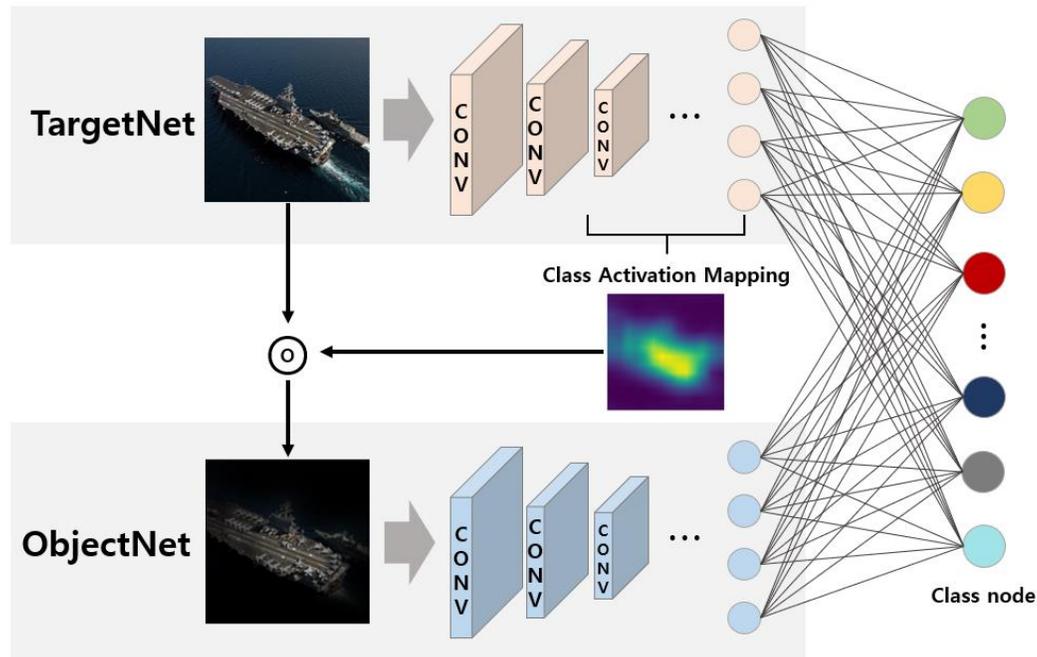


Lakeland terrier

- ImageNet은 1000개의 클래스 분류를 위한 이미지 데이터로서, 수 많은 상위 클래스와 하위 클래스가 존재
- CNN 기반의 모델들이 좋은 성능을 보이며, State-of-art 모델들은 약 85% 수준의 정확도를 보임
- 정확도 개선을 위해 하위 클래스 간의 구별력을 향상 필요, 모델이 하위 클래스 간의 미세한 특징들을 식별해야함
- 이 미세한 특징들은 CNN이 학습하지 못했거나, 배경의 간섭에 의해 특징이 추출되지 않을 수 있음

1. Introduction

관련 연구-Multi CAM



Multicam: Multiple class activation mapping for aircraft recognition in remote sensing images(Remote Sensing, 2019)



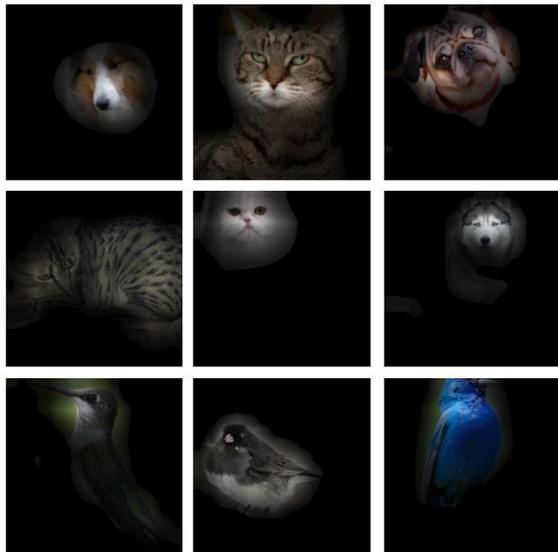
Satellites Aircraft data

Model	Top1-Acc
GoogLeNet	87.75%
GoogLeNet+MultiCAM	88.37%
ResNet	89.80%
ResNet + MultiCAM	91.79%

- TargetNet은 이미지의 특징을 추출하여 분류 및 마스크 이미지 생성
- ObjectNet은 마스크 이미지를 학습하여 이미지의 미세한 특징 추출 및 분류
- Class activation map(CAM) 기반의 마스크 이미지는 배경 간섭을 차단하고, 모델이 미세한 특징을 추출하도록 유도
- Multi CAM은 2개의 CNN을 별도로 학습, 전역 평균 풀링(GAP)이 강제되어 정확도 하락, Pre-trained 모델 사용불가로 범용성이 떨어짐

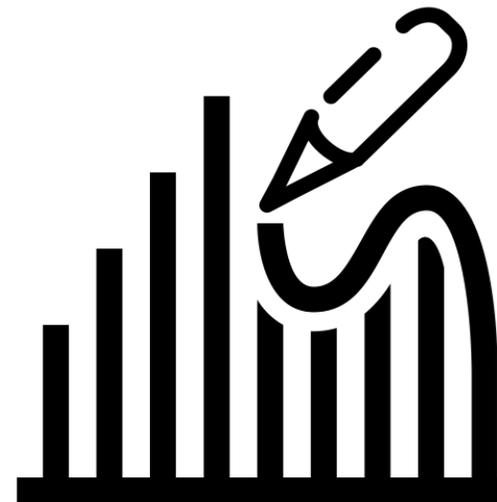
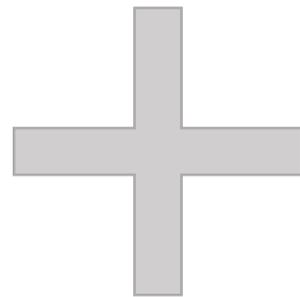
1. Introduction

Concept



마스크 이미지

- 배경 간섭 차단
- 미세한 특징 추출 용이



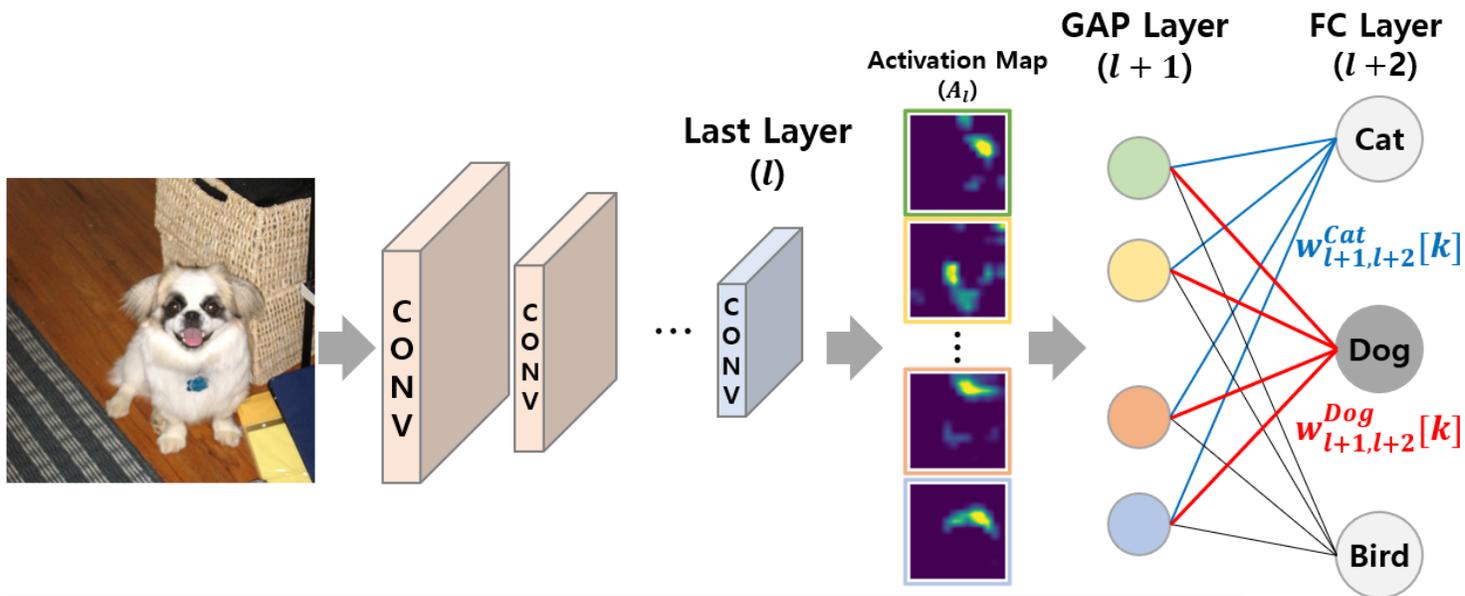
클래스 확률 보정

- 모든 CNN모델 적용 가능
- 분류 성능 향상

마스크 이미지를 이용한 CNN 기반 클래스 확률 보정

2. Background

Class Activation Map(CAM)



Class : Dog

$$w_{l+1,l+2}^{Dog}[1] * \text{Activation Map}_1 + \dots + w_{l+1,l+2}^{Dog}[k] * \text{Activation Map}_k = \text{Attribution Map}$$

Class : Cat

$$w_{l+1,l+2}^{Cat}[1] * \text{Activation Map}_1 + \dots + w_{l+1,l+2}^{Cat}[k] * \text{Activation Map}_k = \text{Attribution Map}$$

Learning deep features for discriminative localization(CVPR, 2016)

- CNN의 추론 결과를 설명하기 위해 이미지에 히트맵 형태로 표현하는 시각적 설명 기법

$$L_{CAM}^c = ReLU(\sum_k \alpha_k^c A_l^k)$$

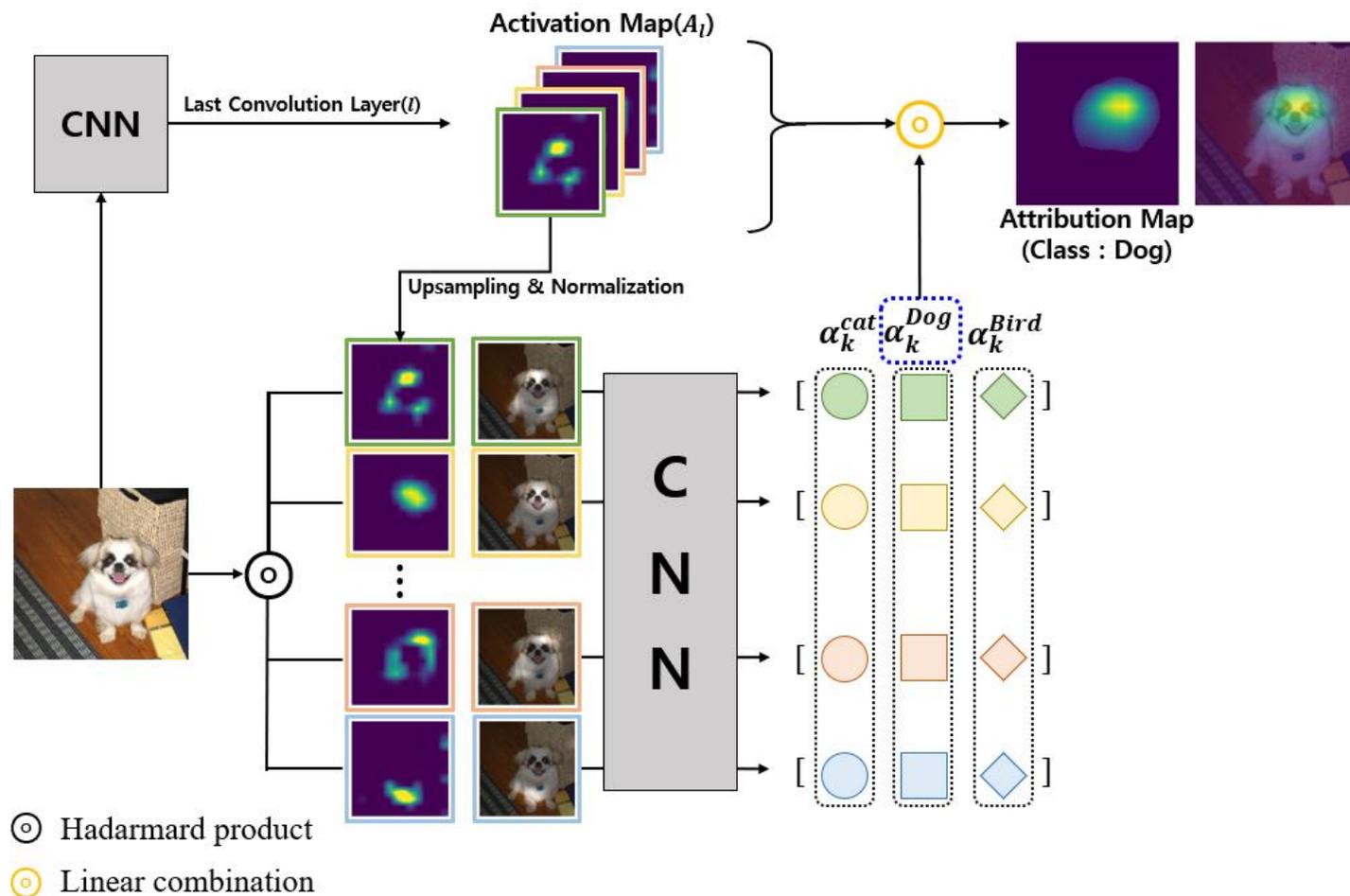
- 활성화맵(A_l^k)과 가중치(α_k^c)의 선형결합을 통해 결과물인 속성 맵(Attribution Map) 생성

$$\alpha_k^c = w_{l+1,l+2}^c[k]$$

- 가중치(α_k^c)는 GAP계층과 FC계층을 연결하는 CNN의 학습 파라미터로 결정
- 클래스에 따라 가중치는 변화하므로 클래스 별 속성 맵 생성 가능

2. Background

Score-CAM



Score-cam: Score-weighted visual explanations for convolutional neural networks(CVPR, 2019)

- CAM과 마찬가지로 활성화 맵과 가중치의 선형결합 개념 유지

$$L_{Score\ CAM}^c = ReLU\left(\sum_k \alpha_k^c A_l^k\right)$$

- 가중치(α_k^c)를 활성화 맵에 의해 변형된 이미지의 클래스 확률로 정의

$$H_l^k = s(UP(A_l^k)) \quad \alpha_k^c = f(X \odot H_l^k)$$

- CAM 기반의 시각적 설명 방법론 중 localization 성능이 뛰어나며, CNN의 추론 과정을 잘 반영함
- CAM과 마찬가지로 입력 이미지에 대하여 모든 분류 클래스의 속성 맵 생성 가능

2. Background

하위 클래스 재정의



Yorkshire Terrier



Black foot ferret



Top1(11.68%)
Yorkshire Terrier



Top2(10.17%)
Black foot ferret

- CNN은 Yorkshire Terrier로 올바르게 분류하였으나, 약 1.5%의 근소한 차이로 Black foot ferret이 산출
- 인간은 두 클래스 간의 연관성을 추론할 수 없지만, CNN은 비슷한 특징을 가진 것으로 판단
- 인간이 판단한 하위 클래스와 CNN이 판단한 하위 클래스 간의 차이가 존재

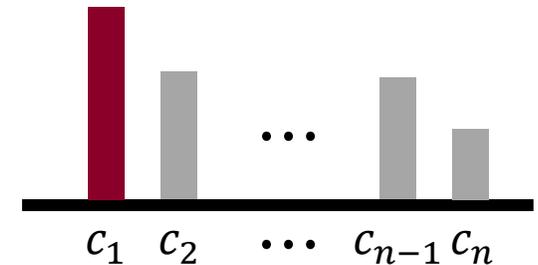
인간에 의해 정의된
하위 클래스
(예 : 개의 품종, 꽃 종류 등)

CNN이 특징을
제대로 구분하지
못하는 클래스

하위 클래스 재정의

$$Y = f(X)$$

※ Y는 입력 이미지에 대한
확률분포
f는 CNN 모델, X는 입력 이미지



- 클래스 확률은 이미지의 하위 클래스를 가장 직관적으로 설명
- Top1 클래스와 근소한 확률차이를 가진 클래스는 하위클래스로 정의할 수 있음
- 하위 클래스는 이미지마다 다르게 인식됨

3. Proposed Method

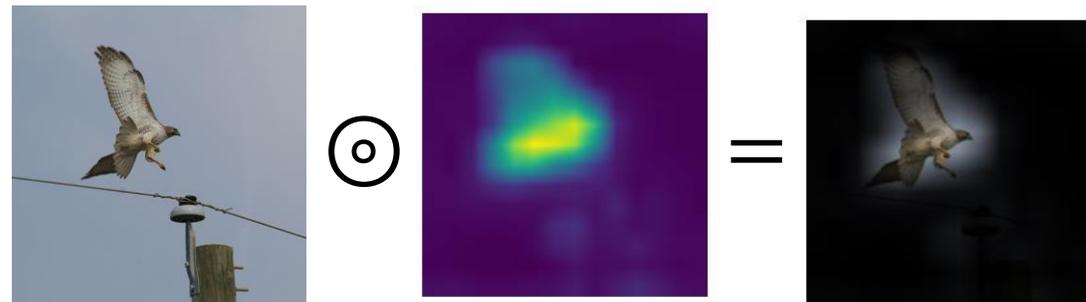
Score-CAM의 활용 및 마스크 이미지 생성 방법



- 이미지는 CNN이 각 클래스로 분류할 때 사용한 중요 구역과 그 외의 불필요한 구역으로 구분
- 각 클래스 별 속성 맵에서 활성화 된 영역은 중요 구역으로 해석 가능
- Score-CAM은 이미지에서 중요 구역과 불필요한 구역을 구분하는 도구로서 활용

Hadamard product, Point-wise manipulation

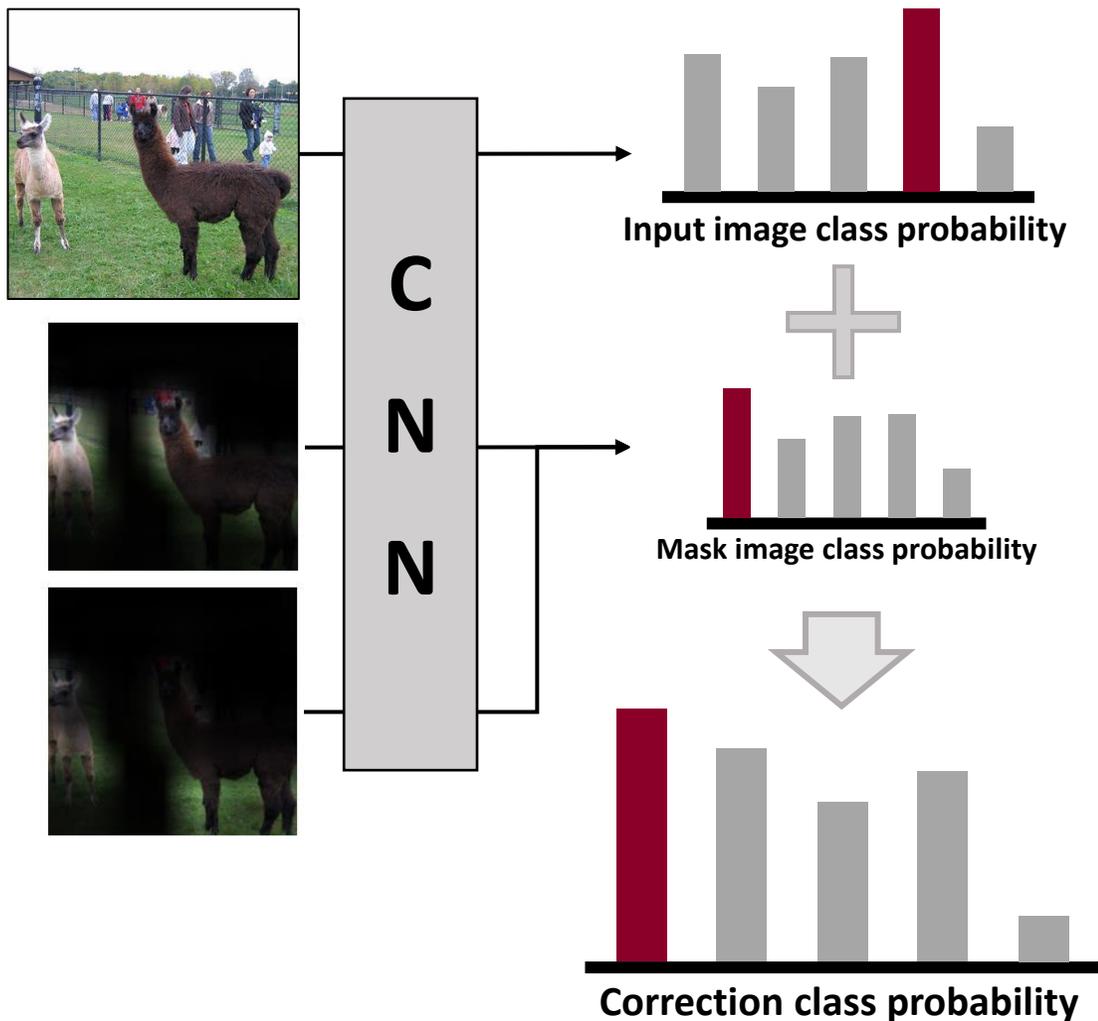
※ 두 이미지의 같은 위치의 원소끼리 곱하는 과정



- Score-CAM의 속성 맵은 픽셀 단위 [0,1]의 값으로 구성
- 입력 이미지의 픽셀값은 Hadamard product를 통해 속성 맵의 값에 따라 강조 되거나 가려짐
- 값이 0일 경우 해당 픽셀값은 완전히 가려지며, 1일 경우 해당 픽셀값은 그대로 유지
- 속성 맵의 0.1이하의 픽셀값은 0으로 치환하여 노이즈 제거

3. Proposed Method

클래스 확률 보정

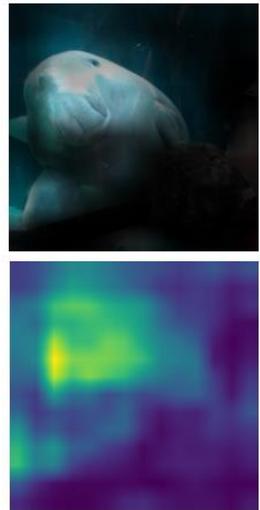
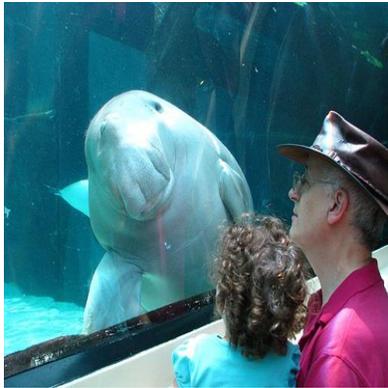


- 마스크 이미지는 CNN이 추론에 활용하지 않은 불필요한 구역의 간섭을 차단하는 역할 수행
- 간섭에 의해 사라졌던 미세한 특징 추출 가능
- 중요 구역의 정확한 클래스 확률 산출 가능
- 입력 이미지를 중요 구역을 활용해 재평가
- 마스크 이미지로 재평가 된 클래스 확률을 보정값으로 정의
- 보정값은 입력 이미지의 클래스 확률에 추가적인 정보 제공

하위 클래스 분류 성능 향상

3. Proposed Method

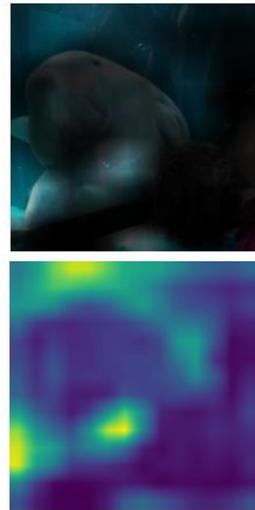
Top k 클래스 선택



Top1
(Dugong)



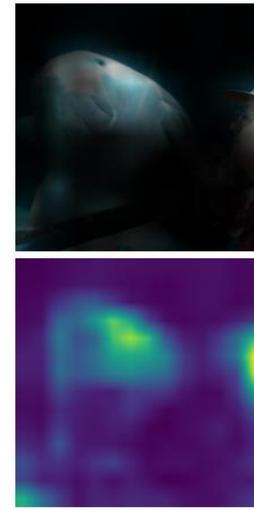
Top2
(Loggerhead turtle)



Top3
(Jelly Fish)



Top4
(Torpedo)

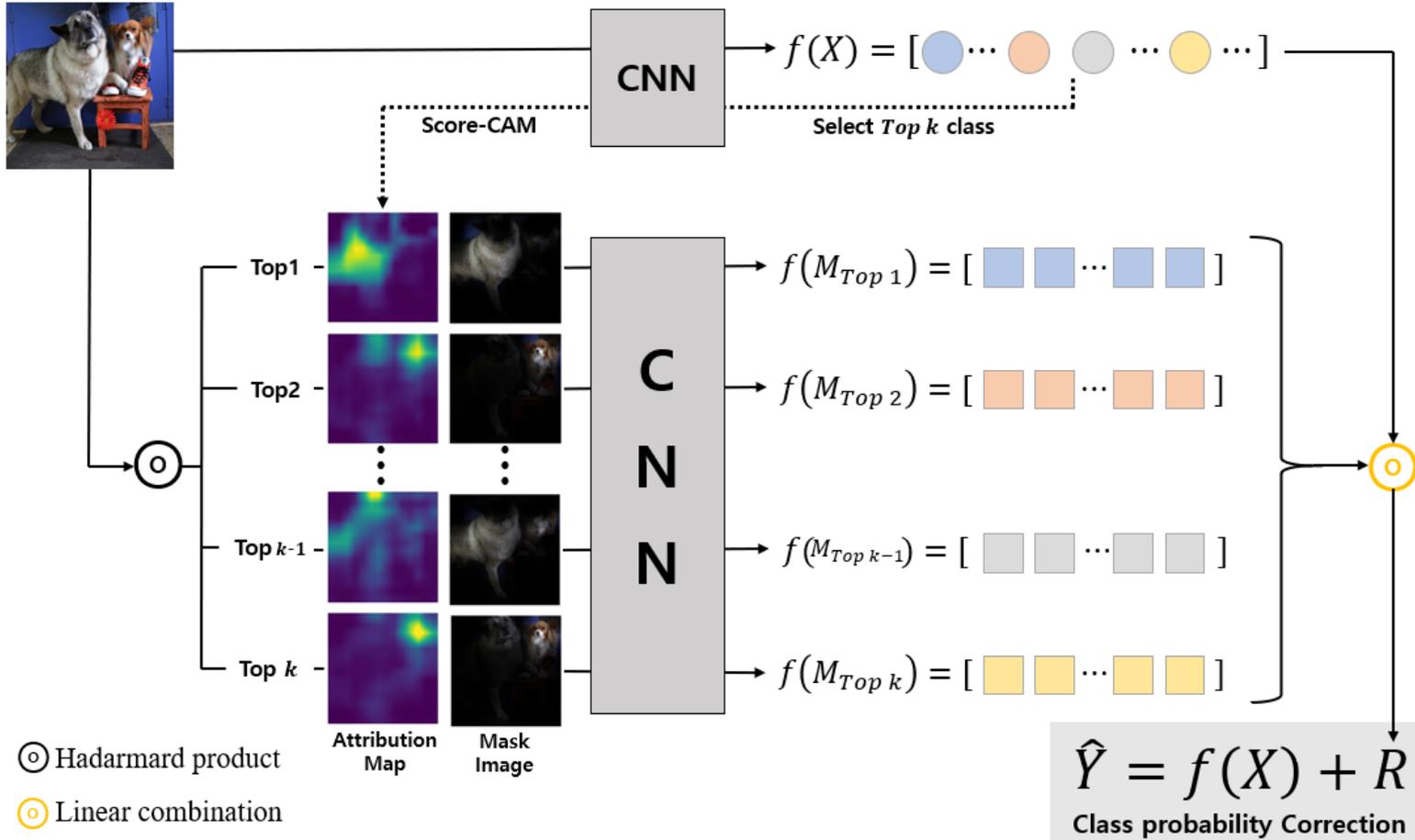


Top5
(White Shark)

- 추가적인 정보를 제공하는 마스크 이미지는 입력 이미지의 Top1부터 Top k 까지 클래스를 선정하여 생성
- 이미지 내의 중요구역을 최대한 찾아내는 목적
- 확률이 낮은 클래스는 CNN의 추론 과정에 큰 영향이 없으며, 보정 효과도 적기 때문에 속성 맵을 생성하지 않음
- 인간이 해당 클래스를 추론하기 힘든 구역도, CNN이 해당 클래스로 추론에 사용한 구역이라면 중요 구역으로 인식

3. Proposed Method

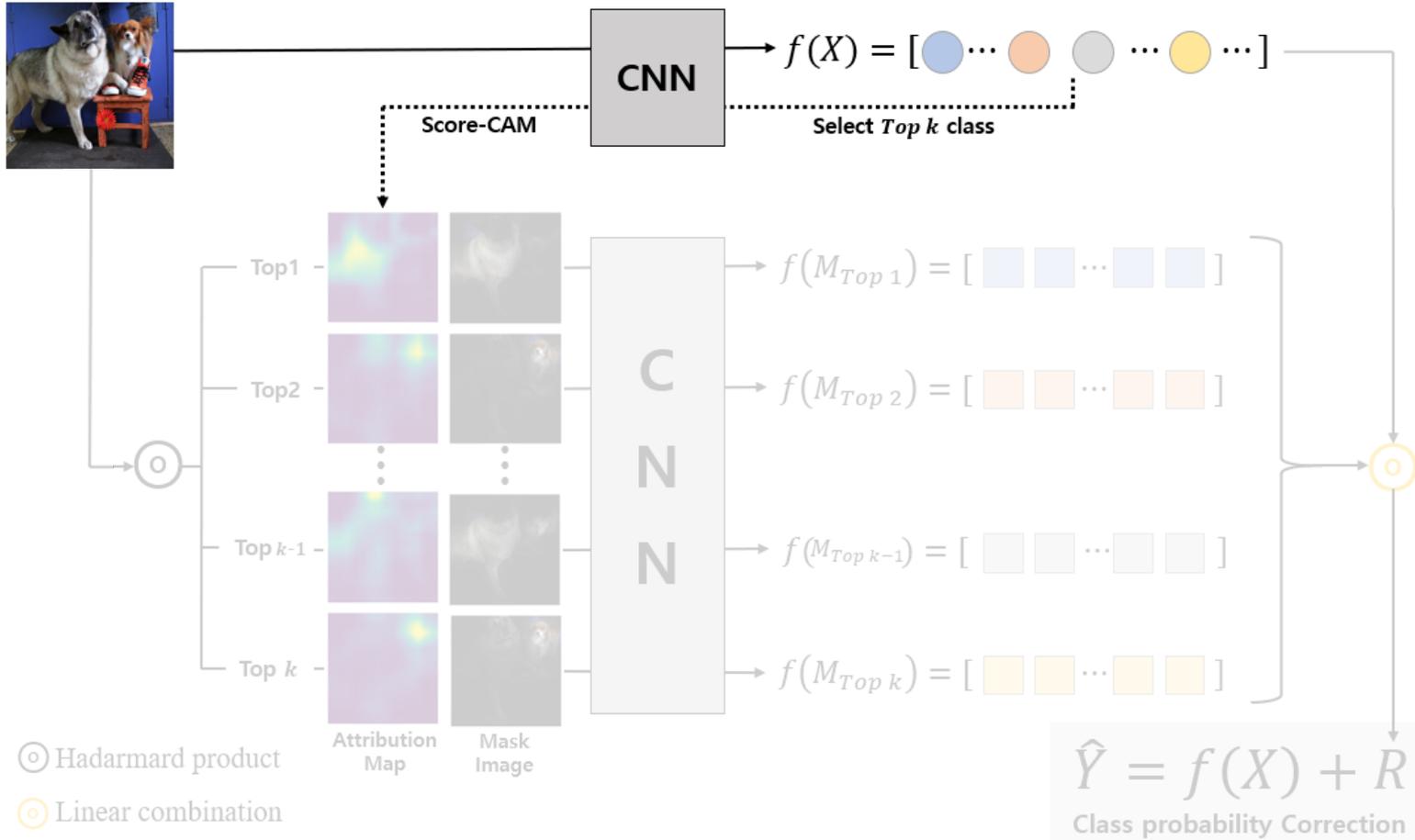
Architecture



- 마스크 이미지를 활용하여 원본 이미지의 클래스 확률을 보정하는 방식 제안
- 클래스 확률 차이가 미세한 경우, 보정값이 더 큰 역할을 할 수 있도록 설계
- 출력값 Y 가 확률 분포인 일반적인 CNN에 모두 적용 가능
- k 는 파라미터로서 $k \in \{1, \dots, k\}$
- Top k 클래스에 대한 마스크 이미지는 $M_{Top\ k}$ 로 정의

3. Proposed Method

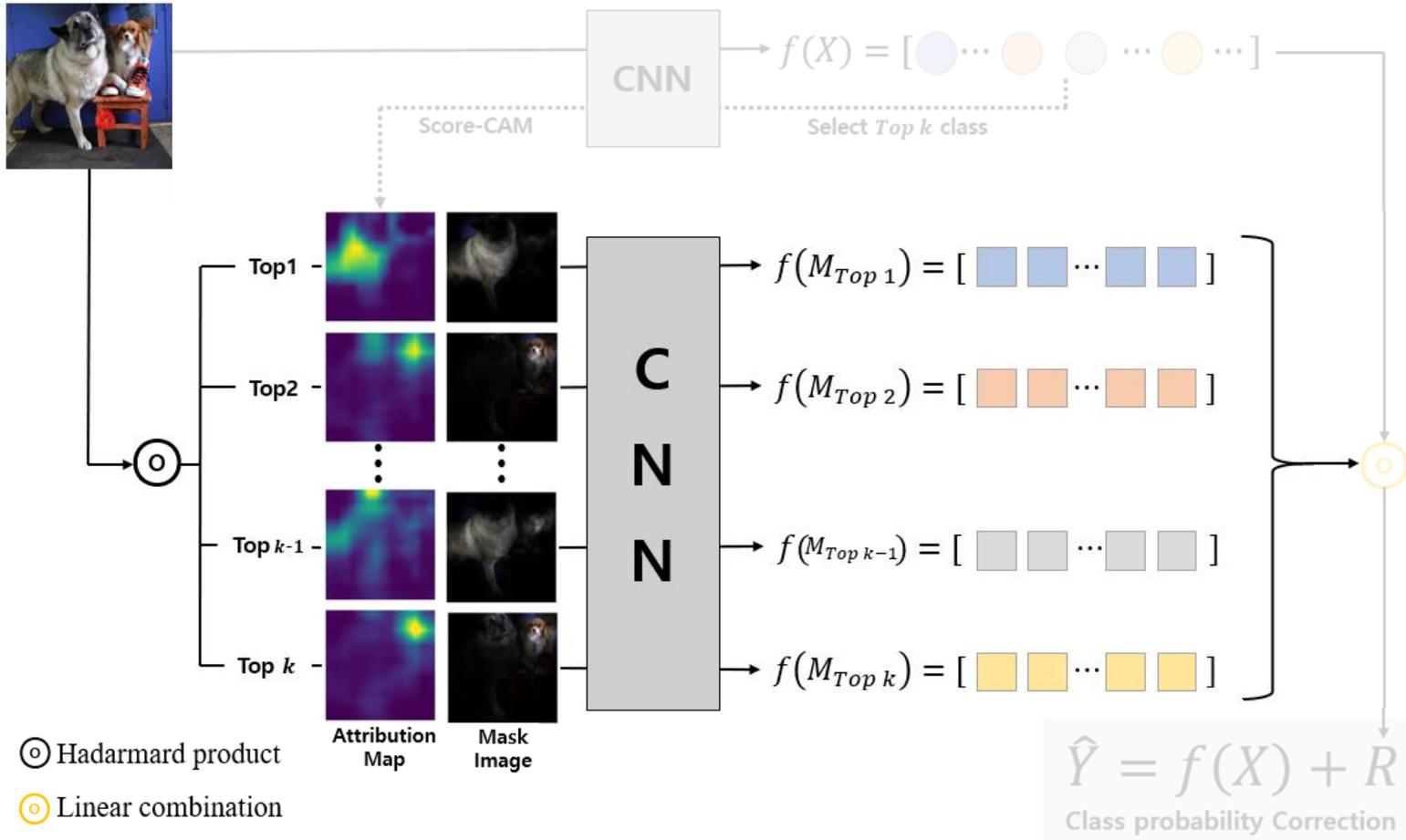
Pipeline



- 입력 이미지를 CNN에 통과시켜 클래스 확률 $Y = f(X)$ 출력
- 출력된 클래스 확률에서 Top k 까지 클래스 선택
- Top 1부터 Top k 까지 Score-CAM으로 속성 맵 생성

3. Proposed Method

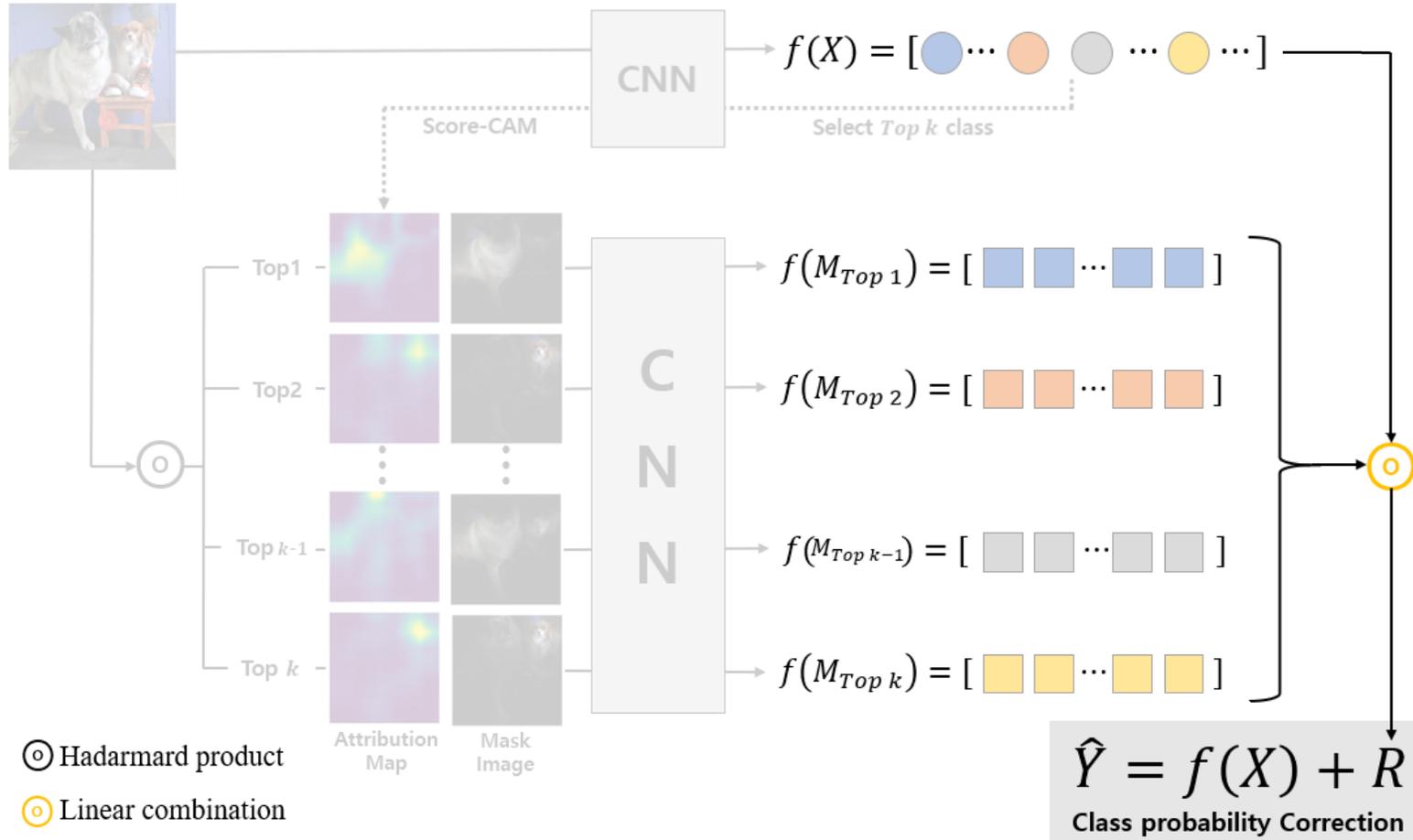
Pipeline



- Top k 까지 속성 맵들과 입력 이미지를 hadarmard product로 마스크 이미지 생성
- 각 마스크 이미지를 CNN에 통과시켜 중요영역에 대한 클래스 확률 $f(M_{Top\ k})$ 산출

3. Proposed Method

Pipeline



- 마스크 이미지의 클래스 확률 산출 후, 보정 진행
- $Y_{Top\ k}$ 는 입력이미지의 Top k 에 해당하는 클래스 확률값(스칼라)
- $f(M_{Top\ k})$ 는 Top k 에 해당하는 마스크 이미지의 클래스 확률(확률분포)
- 보정값 R 은 $f(M_{Top\ k})$ 와 $Y_{Top\ k}$ 의 선형결합

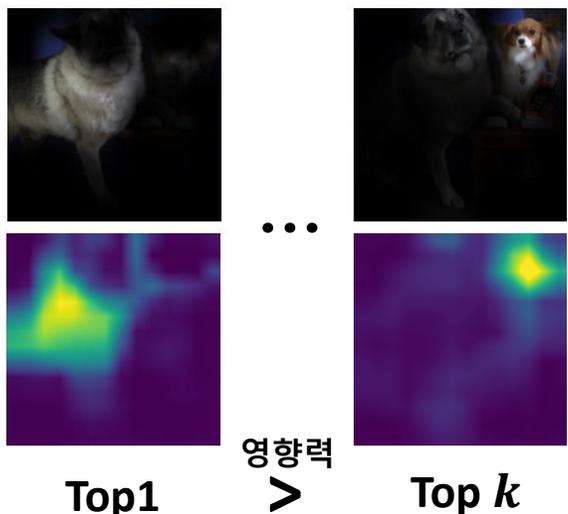
$$R = \sum_1^k f(M_{Top\ k}^c) Y_{Top\ k}^c$$

- 최종적으로 입력이미지의 클래스 확률에 보정값을 더하여 보정된 클래스 확률 \hat{Y} 산출

$$\hat{Y} = f(X) + R$$

3. Proposed Method

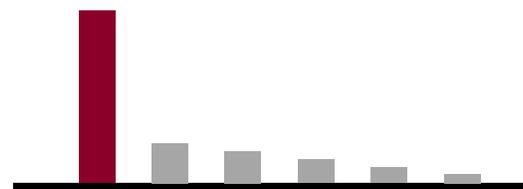
보정 효과



$$R = \sum_1^k f(M_{Top k}^c) \boxed{Y_{Top k}^c}$$

- Top k 에서 Top 1 클래스 순으로 CNN 추론 결과에 더 큰 영향력을 줌
- 보정값 산출 시 Top1 마스크 이미지의 추가 정보는 더 높은 가중치를 가져야함
- 이를 반영하기 위해 입력 이미지의 클래스 확률값들은 마스크 이미지의 클래스 확률 가중치로서 활용

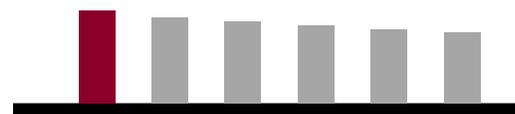
입력 이미지의 클래스 확률 분포



Top1 클래스를 강하게 지지할 경우, 보정 후의 결과가 쉽게 변하지 않음

$$\hat{Y} = f(X) + R$$

영향력 ↑ (blue arrow) ↓ (red arrow)



- 클래스 간 확률 차이가 미세할 경우 보정값의 영향력 상승으로 보정 효과 극대화
- 보정값은 하위 클래스의 미세한 특징 정보가 반영되므로, 더 좋은 구별능력 제공

$$\hat{Y} = f(X) + R$$

영향력 ↓ (red arrow) ↑ (blue arrow)

4. Experiments

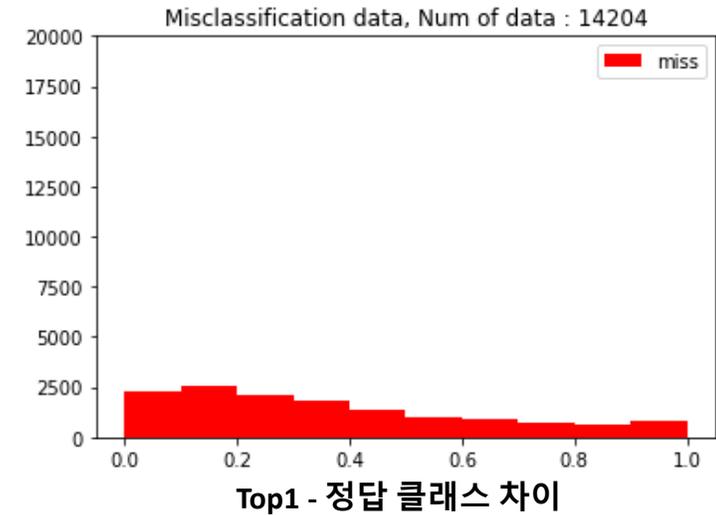
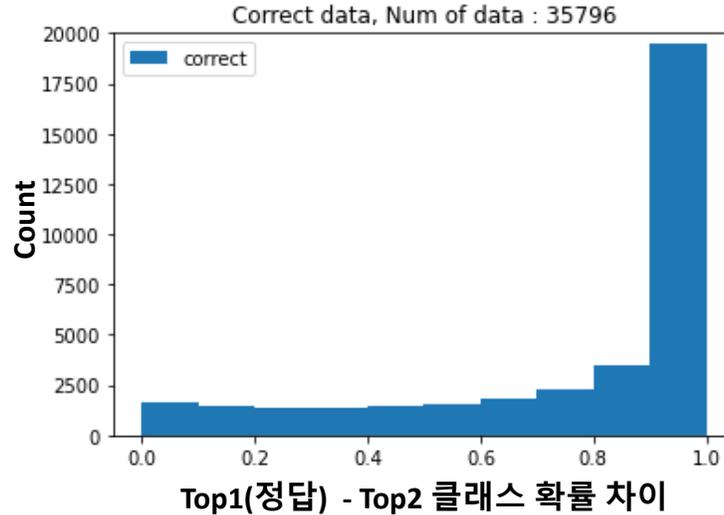
실험 설계

- 클래스 보정의 효과를 검증하기 위한 목적
- ImageNet 데이터를 사용, 테스트 데이터는 총 5만장
- 비교군은 VGG16 모델로서, VGG16의 분류 성능과 VGG16 + 보정의 분류 성능 비교
- VGG16 모델은 Pytorch에서 제공하는 pre-trained 모델 사용
- Top1-정확도만을 고려하였으며, Top5-정확도는 고려하지 않음

4. Experiments

VGG16 Test result

Total	50000
True	35796
False	14204
Top1-Acc.	71.59%
Top5-Acc.	90.38%

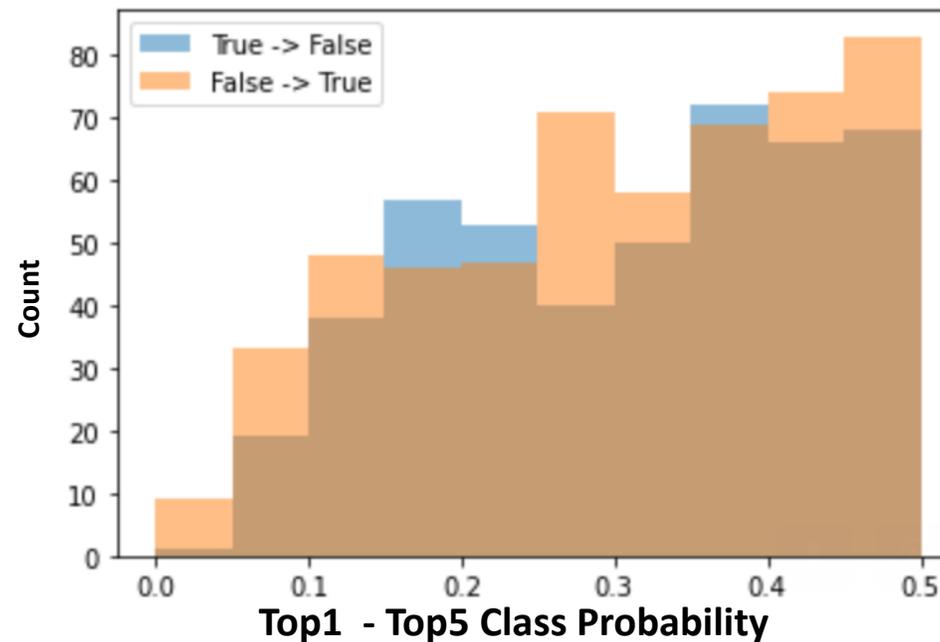


- 정답을 맞춘 이미지는 보정을 통해 예측 결과가 바뀌지 않아야하므로 Top1 - Top2의 확률 차이가 클수록 유리. 확률 차이가 0.5 이하인 데이터가 7272개로 기존에 정답을 맞춘 데이터를 보정을 통해 오분류할 가능성이 존재
- 오분류 이미지는 보정을 통해 예측 결과가 바뀌어야하므로 Top1 - 정답 클래스의 차이가 작을수록 유리. 확률 차이가 0.5 이하인 데이터가 10112개로 보정을 통해 오분류를 정정할 수 있는 가능성이 높음
- 정답 데이터의 히스토그램에서 CNN이 Top1 클래스를 강하게 지지하는 경우, 정답을 맞출 가능성이 굉장히 높으며 클래스 확률을 보정하는 것이 의미가 없을 수 있다는 점을 보임

4. Experiments

Total number of Image : 12491(Top1 – Top5 < 0.5)

VGG16		VGG16 + Correction	
32.1%		32.7%	
VGG16 + Correction			
		True	False
VGG16	True	3553	464
	False	538	7936



- 클래스 확률 보정으로 인한 정확한 성능 평가를 위해 Top1 – Top5 클래스 확률 차이가 0.5 미만인 이미지에 대하여 성능평가
- 조건을 만족하는 이미지들은 Top1부터 Top5까지 클래스 확률 차이가 근소하여 하위클래스를 가진 이미지들로 해석 가능
- 총 12491개의 이미지에 대하여 0.6%(74개)의 성능 향상을 관측
- 제안하는 마스크 이미지를 활용한 클래스 확률 보정으로 하위 클래스에 대한 분류 성능을 향상시킬 수 있다는 것을 증명

5. Conclusion

- 이미지 분류 문제에서 하위 클래스의 구별 능력은 모델의 성능 향상에 중요
- 이를 위해 마스크 이미지를 활용한 CNN의 클래스 확률 보정 방식 제안
- 마스크 이미지는 불필요한 정보의 간섭을 차단하고, 미세한 특징 추출을 유도
- 클래스 확률 보정은 확률 분포를 출력하는 모든 CNN 모델에 적용 가능
- 문제 난이도가 높은 클래스 간 특징이 비슷한 데이터에서 활용 가능할 것으로 기대
- 모델 구조의 재설계, 재학습 및 추가 학습 없이 확률 보정만으로 하위 클래스 분류 성능을 향상시킨 점에서 기여함
- 추후 과제로서 분류 성능 향상이 다양한 데이터와 CNN 모델에서 보였다는 점 증명 필요

감사합니다

맹대주 석사과정, 백준걸 교수*

고려대학교 산업경영공학과

{eownaod, jungeol}@korea.ac.kr