💙

[5주차] Relevance-CAM: Your Model Already Knows Where to Look

0. Abstract

- Relevance - CAM
  - 새로운 CAM 방법: Relevance-CAM은 Layer-wise Relevance Propagation을 활용
  - 중간 레이어 분석 기능: 마지막 컨볼루션 레이어뿐만 아니라 중간 레이어도 분석 가능
- Relevance-CAM의 장점
  - shattered gradient 문제에 강건, 설명 맵이 신뢰할 수 있음
  - 어떤 깊이의 레이어에서도 인식 및 위치 평가에서 다른 방법보다 우수함

1. Introduction

- 모델의 결정 과정을 이해하고 해석하는 방법에 대한 연구의 중요성
  - 의료 영상 처리와 같은 중요한 분야에서 모델의 결정을 정확히 이해하는 것이 매우 중요
  - 이를 위해 Class Activation Map(CAM)과 같은 기존 방법들이 사용되고 있지만, 모델 구조에 제약이 있거나 특정 레이어에만 초점을 맞추는 한계가 존재
- Relevance-CAM
  - 모델의 어느 레이어에서든(특히 중간 레이어에서도) 잘 작동함
  - 모델이 어떻게 결정을 내리는지에 대한 더 자세한 분석을 가능하게 해줌
  - shattered gradient 문제에 강하며, 얕은 레이어에서도 클래스별 특징을 추출할 수 있다는 놀라운 사실이 발견됨

- 작동 과정
  - 히트맵을 사용하여 모델의 결정 과정을 시각화, 모델이 어떤 정보를 중요하게 여기는지 보여줌
  - 기존의 CAM 기반 방법들보다 목표 객체를 더 정확하게 지역화할 수 있으며, 특히 얕은 레이어에서의 성능이 뛰어남
    - **지역화**: 모델이 이미지 내에서 특정 객체를 식별하고, 그 위치를 정확하게 찾아내는 과정
    - CAM은 모델이 특정 클래스를 식별하기 위해 **이미지의 어느 부분에 주목하는지** 보여주는 히트맵을 생성함
- 평가 지표
  - Average Drop, Average Increase, Intersection over Union과 같은 지표를 통해 객관적으로 평가됨
  - 중간 레이어에서 특히 우수한 성능을 보임
  - Relevance-CAM이 딥러닝 모델의 해석 가능성을 한 단계 끌어올리는 중요한 기여를 하고 있음을 보여줌

➡️ **Relevance-CAM은 딥러닝 모델의 결정 과정을 더 잘 이해하고 해석할 수 있는 새로운 방법을 제공하며, 특히 중요한 의사 결정이 필요한 분야에서 모델을 더 안전하고 효과적으로 사용할 수 있도록 함**

2. Background

2.1 CAM

Class Activation Mappeing(CAM)

- global 풀링 레이어 이전의 마지막 컨볼루션 레이어 출력의 선형 가중치 조합을 통해 클래스 특정 영역을 시각화하는 설명 방법
- 특정 아키텍처를 가진 모델에만 적용이 제한됨

2.2 Grad-CAM

Grad-CAM

- 딥러닝, 특히 CNN에서 어떤 영역이 특정 클래스를 예측하는 데 중요한지 시각화하는 기술
- CAM을 일반화하여 모든 CNN 모델에 적용할 수 있도록 설계됨

- 수식

$$L^c_{Grad-CAM} = \sum_k \alpha^c_k A_k$$

  - A_k: 마지막 컨볼루션 레이어의 k번째 채널에서의 활성화 맵, 이미지의 어떤 부분이 중요한지를 나타냄
  - y^c: 클래스 c에 대한 모델 출력
  - (α_k)^c: Grad-CAM의 가중치 구성 요소로, 각 활성화 맵의 중요도를 나타냄

$$\alpha^c_k = GP(\frac{\partial y^c}{\partial A_k})$$

  - 가중치 (α_k)^c
    - 클래스 c에 대한 활성화 맵 (A_k)의 그래디언트를 글로벌 평균 풀링하여 계산
    - GP: 글로벌 풀링 함수
    - 각 활성화 맵이 특정 클래스를 예측하는 데 얼마나 중요한지 나타내는 계산

2.3 Layer-wise Relevance Propagation(LPR)

Layer-wise Relevance Propagation(LPR)

- 모델의 결정 과정을 이해하기 위해 개발된 기술
- 모델이 특징 결정을 내릴 때 어떤 입력 픽셀이 중요했는지를 역추적하여 시각화
- 모델의 '블랙박스'적인 성격을 해소할 수 있음
- 기본 원리
  - 모델의 출력부터 시작하여 입력층까지 역으로 **관련성 점수**(relevance score)를 전파
  - 관련성 점수는 모델의 결정에 각 입력 픽셀이 얼마나 기여했는지를 나타냄
- 보존성(Conservativeness)

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

  - 모델이 감지한 **총 관련성 점수**는 **입력 픽셀에 할당된 관련성 점수의 합**과 동일
  - f_c(x): 타겟 클래스 c에 대한 모델의 출력 점수
  - (R_p)^l(x): l번째 레이어에서 픽셀 p에 할당된 관련성 점수
- 양성성(positivity)

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

  - 히트맵을 형성하는 모든 값(모든 관련성 점수)은 0 이상
- z-rule

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

$$z_{ij}^+ = x_i w_{ij}^+$$

  - Deep Taylor Decomposition에 기반한 전파 규칙
    - 가중치의 양의 부분만을 고려하여 긍정적인 기여도만을 전파
  - j번째 레이어의 타당성 점수 R_j를 i번째 레이어로 전파하는 방법에 대한 공식
    - R_i, R_j: i, j번째 레이어의 관련성 점수
    - x_i: i번째 레이어의 활성화 출력
    - (w_ij)^+: i와 j 레이어 사이의 가중치의 양의 부분

- 의의와 한계
  - 의의: 모델이 특정 결정을 내릴 때 중요하게 고려한 입력 영역을 시각적으로 파악 가능 ➡️ **모델의 해석 가능성을 높이는 데 큰 도움이 됨**
  - 한계: z-rule은 다중 객체 이미지에서 타겟 클래스에 대한 민감도가 떨어짐

2.4 Contrastive Layer-wise Relevance Propagation(CLRP)

Contrastive Layer-wise Relevance Propagation(CLRP)

- 타겟 클래스에 대한 관련성 점수에서 비타겟 클래스의 관련성 점수를 빼는 방식
- 타겟 클래스에 대한 히트맵의 민감도를 더욱 높이며 이로 인해 생성된 히트맵은 타겟 클래스에 대해 더욱 민감하게 반응함
- 관련성 점수

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

  - n=t일 때: 타겟 클래스에 대한 관련성 점수
    - (z_t)^(L): L번째 레이어에서 타겟 클래스 인덱스 t에 대한 모델 출력값
    - 타겟 클래스에 대한 관련성 점수 = 해당 클래스의 모델 출력값 그 자체
  - otherwise: 비타겟 클래스에 대한 관련성 점수
    - N: 클래스의 총 개수, N-1은 타겟 클래스를 제외한 나머지 클래스의 수
    - 비타겟 클래스에 대한 관련성 점수 = 타겟 클래스의 모델 출력값을 (N-1)로 나눈 후 음수로 만든 값 ➡️ **비타겟 클래스의 중요도를 감소시키기 위함**
    - 비타겟 클래스의 관련 픽셀은 생성된 saliency map에서 제거됨
- CLRP의 중요성
  - 모델이 특정 클래스를 식별하는 데 있어 어떤 픽셀이 중요한지 더 명확하게 보여줌

- 모델의 결정 과정을 이해하고, 모델의 해석 가능성을 높이는 데 큰 도움이 됨
- 타겟 클래스에 대한 민감도를 높임으로써, 모델이 왜 특정 결정을 내렸는지에 대한 더 정확한 설명 제공 가능

2.5 Gradient Issue

1️⃣ Noisiness and discontinuity

- 문제: 그래디언트가 신경망이 깊어짐에 따라 노이즈가 많고 불연속적이 됨
- 원인: 활성화 함수를 통과할 때 그래디언트가 포화되어 사라지거나 폭발적으로 증가함
- 예시: ReLU나 Sigmoid와 같은 활성화 함수를 통과할 때, 그래디언트가 포화되어 사라지거나 폭발적으로 증가함
- 결과: 그래디언트의 조각별 선형성으로 인해 불연속성이 발생하며, 연결된 레이어의 가중치에 의해 계산되기 때문에 다음 픽셀과의 관계가 손실됨

2️⃣ Explanation to Sensitivity

- 문제: Grad-CAM은 활성화 맵의 중요도를 출력에 대한 활성화 맵의 그래디언트로 측정함
- 한계: 활성화 값 자체를 고려하지 않고 중요도를 할당함
- 결과: 활성화 맵의 모델 출력에 대한 민감도만을 측정하며, 실제로는 활성화 맵이 타겟 클래스 출력에 **얼마나 기여하는지**를 설명하는 것이 아니라 **얼마나 민감한지**를 나타냄

3️⃣ Relevance Score

- 장점: LRP는 그래디언트 문제에 강건함이 실험을 통해 입증됨
- 특징: 타겟 클래스 점수가 변할 때 연속성을 보이고 노이즈가 덜하다는 특성을 보임
- 이점: LRP의 관련성 점수는 클래스 활성화 매핑의 가중치 구성 요소로 고려됨

3. Relevance-weighted Class Activation Map

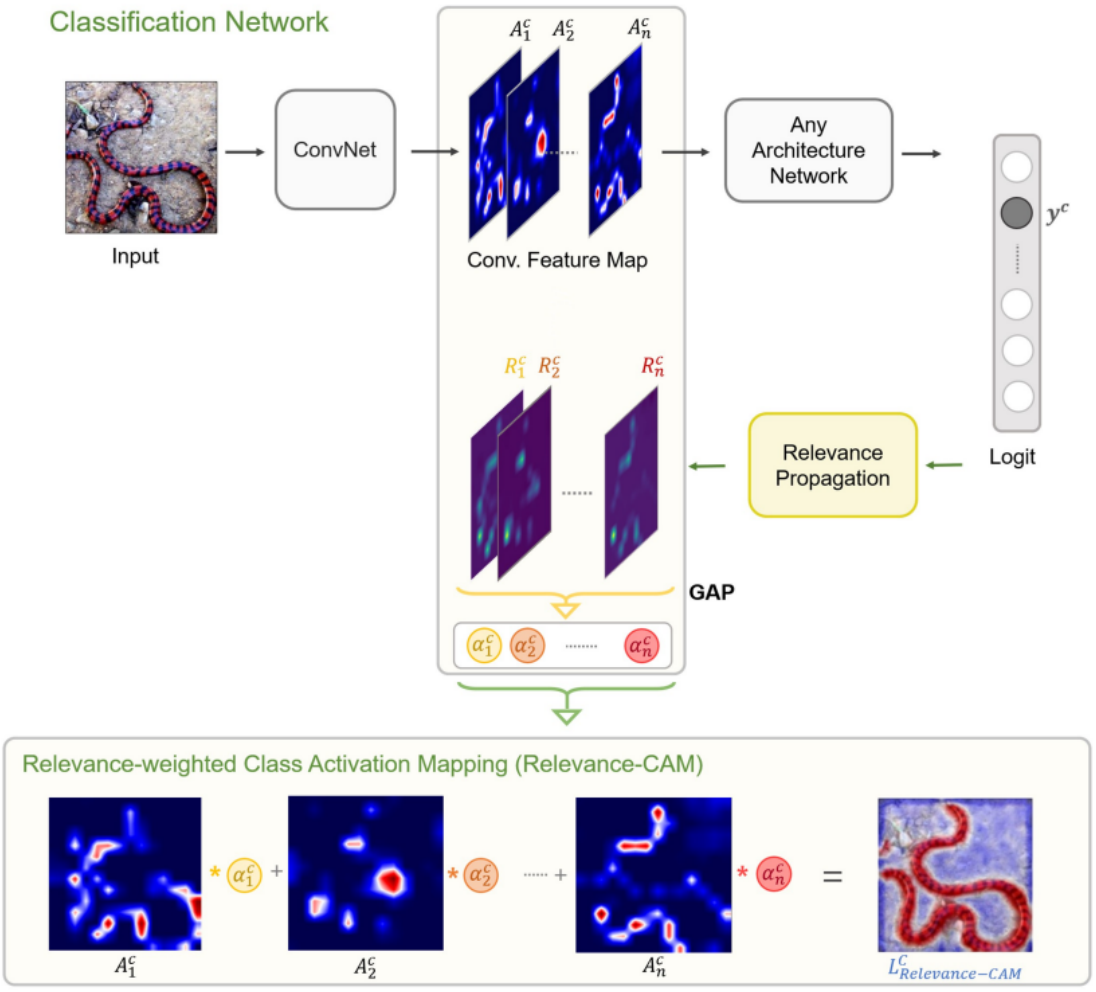

Figure 2. Relevance-CAM pipeline. Activation maps, A_k^c, are extracted during forward propagation and Relevance Maps, $R^{(c,i)}$, are calculated by Relevance propagation process. And the weighting components are obtained by global average pooling of relevance map. Finally, Relevance-CAM is obtained by weighted linear summation of activation maps

Relevance-CAM의 도입 배경

- 이전 방법의 한계: 중간 레이어들의 히트맵이 노이즈가 많고 클래스 특정적이지 않은 결과를 보여주었음
- 새로운 접근: Relevance-CAM은 **중간 레이어들도 클래스 특정 정보를 추출할 수 있음**을 발견

Relevance-CAM의 원리

- 그래디언트 문제: 깊은 레이어 모델에서 그래디언트가 노이즈가 많고 불연속적이 되면서, 그래디언트를 가중치 구성 요소로 사용하는 것의 질이 의문시됨
- Relevance 점수: LRP를 통해 얻은 관련성 점수를 가중치 구성 요소로 사용함

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

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

- (L_(Relevance-CAM)^(c,i)): 클래스 c에 대한 i번째 레이어의 Relevance-CAM
- 모든 채널 k에 대해 합산
- (α_k)^(c,i): k번째 채널의 가중치 구성 요소, 해당 채널이 클래스 c의 출력에 기여하는 정도를 나타냄
  - 전역 평균 풀링을 통해 계산됨
  - 모든 위치(x, y)에 대한 합산
  - (R_k)^(c,i)(x,y): LRP 과정을 통해 얻은 i번째 레이어의 k번째 채널에서 위치 x, y에 대한 관련성 맵
- (A_k)^c: k번째 채널의 활성화 맵

➡️ **LRP를 통해 얻은 관련성 맵을 사용하여 각 채널의 가중치를 계산하고, 이를 통해 클래스에 대한 기여도를 평가**

➡️ **Relevance-CAM은 단 한 번의 전파와 역전파만으로 중간 레이어에서도 클래스 특정 정보를 추출할 수 있도록 함**

Relevance-CAM의 의의

- **중간 레이어의 중요성**: Relevance-CAM을 통해 중간 레이어들도 마지막 컨볼루션 레이어만큼이나 클래스 특정 정보를 추출할 수 있음을 확인함
- **단일 전파 및 역전파로 계산 가능**: 활성화 맵과 관련성 맵을 통해 Relevance-CAM을 계산할 수 있으며, 이는 단 한 번의 전파와 역전파만으로 가능함

4. Experiment

실험 방법(후행 주의 방법을 평가)

- 심층 비교: 다양한 CAM 기반 방법으로 생성된 깊이별 히트맵을 시각화하여 Relevance-CAM의 효과성을 비교함
- 신뢰도 측정: 생성된 히트맵의 신뢰도를 평가하기 위해 Average Drop(A.D.)과 Average Increase(A.I.)를 측정함
- Grad-CAM과의 차이점: Grad-CAM과 Relevance-CAM 사이의 차이점을 보여줌
- 위치 정확도 평가: 교차 오버 유니온(Intersection over Union, IoU)을 통해 위치 정확도를 평가함
- 클래스 민감도 평가: 제안된 방법의 클래스 민감도를 평가함

4.1 Depth-wise visualization

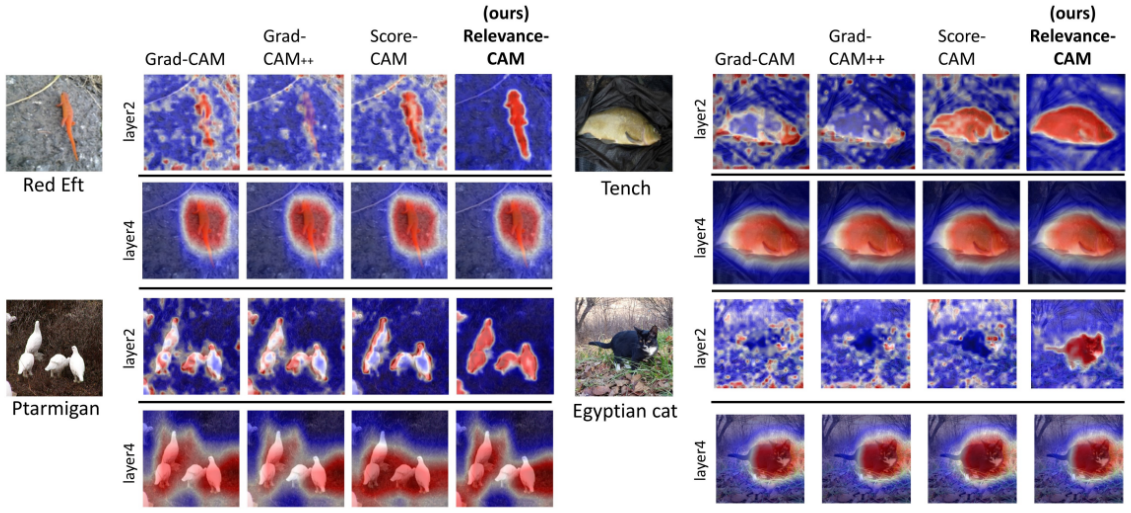

Figure 3. Comparison of various methods. The columns are divided by the explanation methods. The rows are divided along the layer depth. layer 2 is the intermediate layer and layer 4 is the last convolutional layer. In the deep layer, layer4, the heatmaps are similar for the various methods. But in the shallow layer, layer2, the quality of Relevance-CAM is better than that of the other methods in localizing the target objects. And Relevance-CAM shows high resolution heatmaps at low level layer.

	Layer2		Layer4	
Method	A.D.	A.I.	A.D.	A.I.
Grad-CAM	74.91	4.45	23.13	24.05
Grad-CAM++	71.15	4.85	22.03	25.35
Score-CAM	56.59	8.8	21.89	24.65
Relelvance-CAM	**39.02**	**16.6**	**21.53**	**25.7**

Table 1. Lower Average Drop(A.D.) and higher Average Increase(A.I.) indicate better performance. Evaluation for ResNet 50. Layer 2 is the low level layer and, Layer 4 is the last convolutional layer.

	layer23		layer43	
Method	A.D.	A.I.	A.D.	A.I.
Grad-CAM	85.43	1.5	23.15	22.35
Grad-CAM++	86.77	1.3	22.98	22.35
Score-CAM	76.11	3.2	**21.22**	23.5
Relelvance-CAM	**72.25**	**3.9**	22.42	**24.95**

Table 2. Lower Average Drop(A.D.) and higher Average Increase(A.I.) indicate better performance. Evaluation for VGG 16 with batch normalization. Layer 23 is the 3-th maxpooling layer and layer 43 is the 5-th and last maxpooling layer.

주요 발견

- 레이어 4에서의 성능: 모든 설명 방법이 레이어 4에서 대상 객체를 잘 지역화함

- 레이어 2에서의 차이: 그라디언트 기반 방법(예: Grad-CAM, Grad-CAM++)은 레이어 2에서 클래스 특정 영역을 지역화하지 못하지만, Score-CAM과 Relevance-CAM은 잘 지역화함

그라디언트 문제와 Relevance-CAM의 강점

- 그라디언트 기반 방법의 한계: 그라디언트가 레이어를 통과하면서 노이즈가 발생, 활성화 맵에 정확한 가중치를 할당하지 못함
- Relevance-CAM의 강점: 그라디언트 문제에 강한 Relevance-CAM은 얕은 레이어에서도 잘 작동하며 Relevance-CAM의 히트맵이 Score-CAM보다 더 명확함

평가 방법

- 목표 충실도 평가: Average Drop(A.D.)과 Average Increase(A.I.) 메트릭을 사용하여 어텐션 맵이 모델을 얼마나 잘 설명하는지 평가함
  - Average Drop: 어텐션 맵을 사용하여 이미지의 일부를 마스킹했을 때, 모델이 해당 클래스에 대해 예측하는 점수가 얼마나 감소하는지를 측정 ➡️ **모델이 어떤 영역을 중요하게 생각하는지를 이해할 수 있도록 함**
  - Average Increase: 어텐션 맵을 사용하여 이미지의 일부를 마스킹했을 때, 모델이 해당 클래스에 대해 예측하는 점수가 얼마나 증가하는 경우의 비율을 측정 ➡️ **어텐션 맵이 모델의 예측에 어떤 영향을 미치는지를 평가하도록 함**
- 실험 결과: ResNet 50과 VGG16 모두에서 깊은 레이어에서는 A.D.와 A.I. 값이 비슷하지만, 낮은 레이어에서는 각 설명 방법 간의 평가 값 차이가 증가함

4.2 Evaluation for selectivity

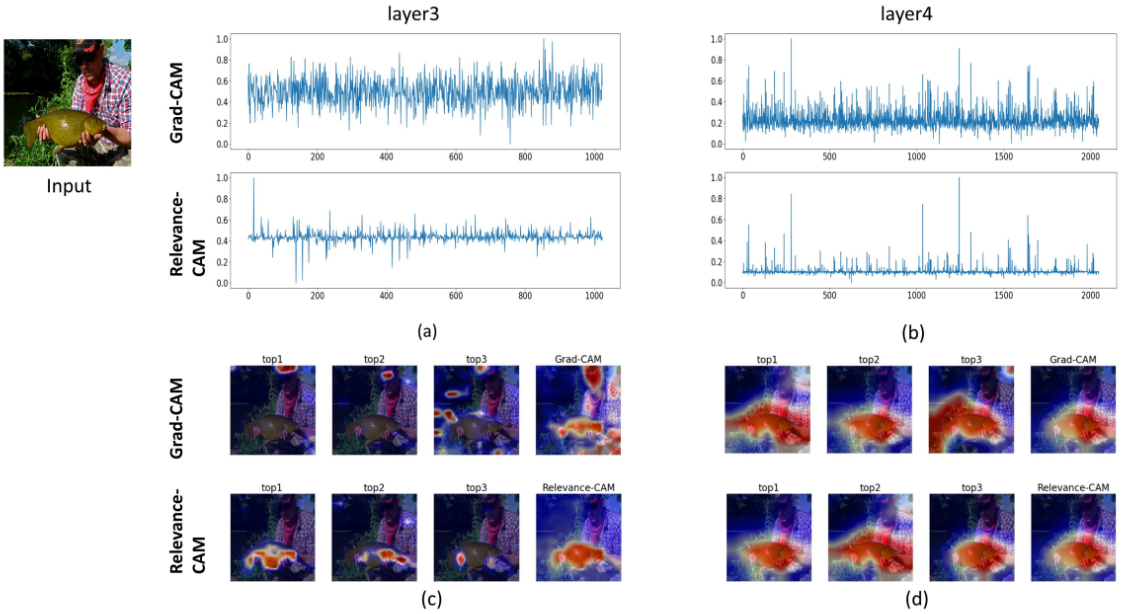

Figure 4. Evaluation of the selectivity. First column and second column show the results for layer3 and layer4 respectively. (a), (b) show the weighting component for each method. (c), (d) show the activation maps of the top 3 weighted channel for each method and the generated attention maps at the end.

채널별 가중치 구성 요소 시각화

- 채널별 가중치 비교: Figure 4의 (a), (b)는 각 방법에 따른 채널별 가중치 구성 요소를 보여주며, 여기서 가중치 구성 요소는 일정한 스케일로 정규화됨
- Shattered Gradient 문제: 3번 레이어에서 gradient 가중치는 노이즈가 많고 평탄화되어 shattered gradient 문제가 발생하는 반면, Relevance-CAM의 관련성 가중치는 3번과 4번 레이어 모두에서 희소성을 보여주며 해당 문제에 있어서 강건함을 입증함

Grad-CAM과 Relevance-CAM의 선택성 비교

- 상위 3개 가중치 활성화 맵: 그림 4의 (c), (d)에서는 Grad-CAM과 Relevance-CAM의 상위 3개 가중치 활성화 맵과 생성된 살리언시 맵을 보여줌
- 레이어 4에서의 성능: 두 방법 모두 레이어 4에서 타겟 클래스 객체(텐치)를 잘 지역화하며, 상위 3개 가중치 활성화 맵은 타겟 객체를 높은 해상도로 강조함
- 레이어 3에서의 차이: 레이어 3에서 Grad-CAM은 노이즈가 많은 히트맵을 보여주며 가장 가중치가 높은 활성화 맵조차 타겟 클래스 특징을 추출하지 못하는데, 이는 Grad-CAM의 노이즈가 많은 가중치 구성 요소가 중요한 특징 맵을 선택하지 못함을 의미함. 반면 Relevance-CAM은 텐치를 명확하게 지역화하며, 상위 3개 가중치 활성화 맵도 타겟 객체를 강조하는 모습을 확인할 수 있음

4.3 Evaluation for Localization

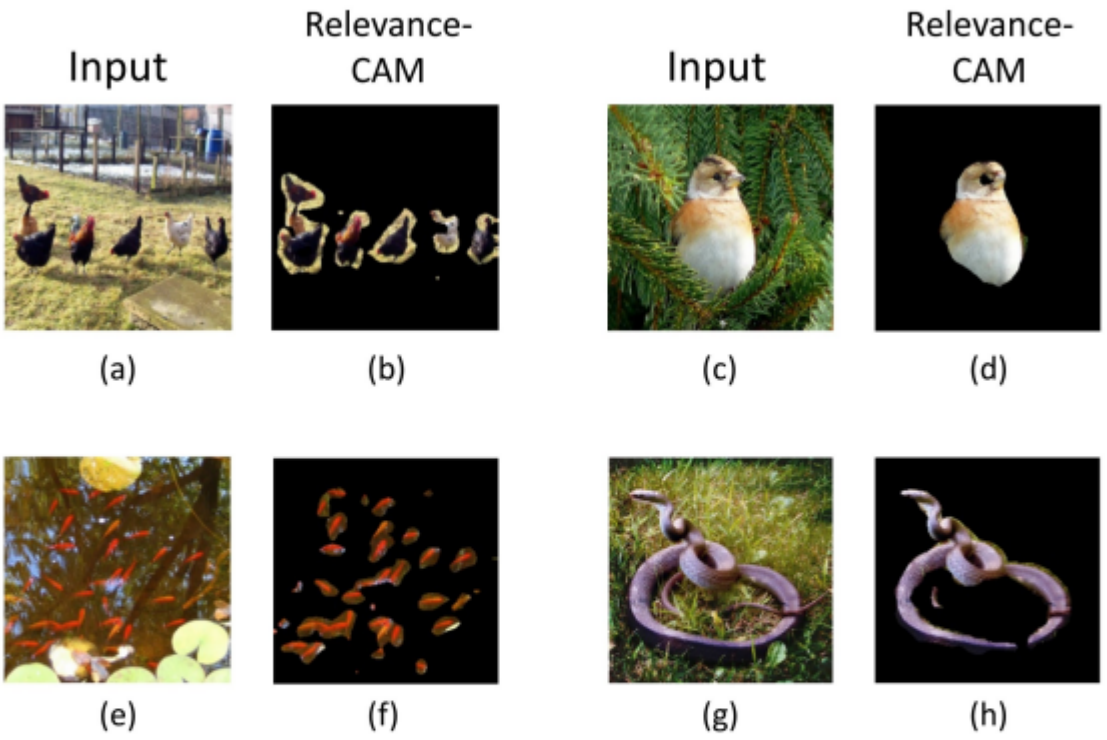

Figure 5. Visualization of weakly supervised localization. (a), (c), (e), and (f) are images for hen, brambling bird, gold fish and ring-neck snake, respectively. (b), (d), (f), and (h) are segmentation image through Relevance-CAM

Relevance-CAM을 이용한 세분화 실험

- 실험 설정: Relevance-CAM을 사용하여 마스크를 적용한 세분화가 수행되며, 이 마스크는 생성된 살리언시 맵의 평균 + 1*(표준 편차)보다 높은 픽셀로 구성됨.
- 결과: Relevance-CAM은 객체를 배경에서 잘 분리하며, 작은 객체조차도 높은 정밀도로 지역화함
    - 금붕어와 같은 작은 객체도 정확하게 지역화

정량적 위치 결정 능력 평가

- 성능 평가 지표: 성능은 Intersection over Union (IoU) 메트릭을 사용하여 평가됨
    - IoU: 생성된 주의 맵의 픽셀과 경계 상자 픽셀 사이의 교차 영역과 합집합 영역의 비율

    - 살리언시 맵이 대상 객체를 밀접하게 지역화하면 IoU 값이 높아짐
- 실험 결과: 모델이 정확하게 예측한 무작위로 선택된 2000개 이미지에서 실험이 수행됨
    - 마지막 컨볼루션 레이어에서는 거의 비슷한 성능 결과가 나타남
    - 얕은 레이어에서는 Relevance-CAM이 다른 방법들을 능가하며, 대상 레이어가 얕아질 때도 Relevance-CAM의 IoU 변화가 적다는 점이 주목됨

4.4 Class Sensitivity Test

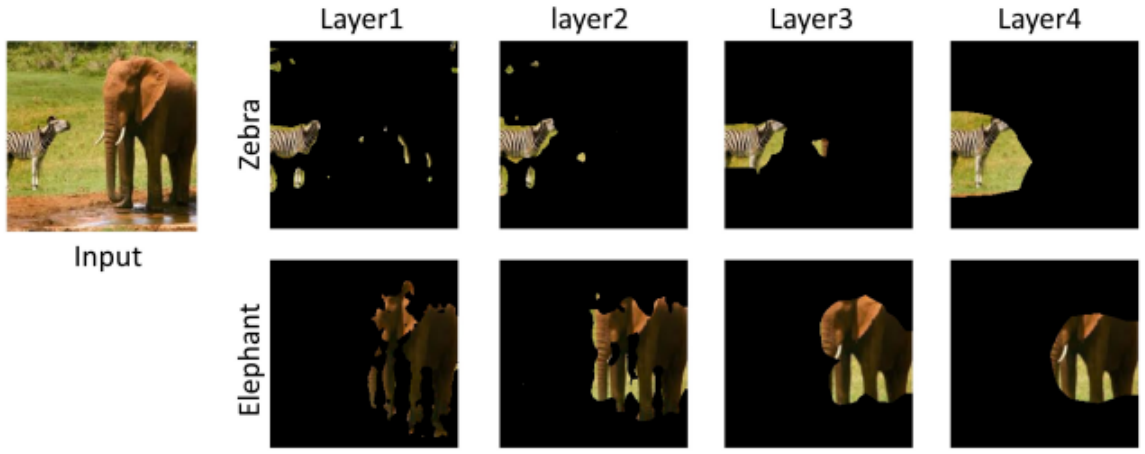

Figure 6. Class sensitivity test of Relevance-CAM on ResNet 50 for multi objects

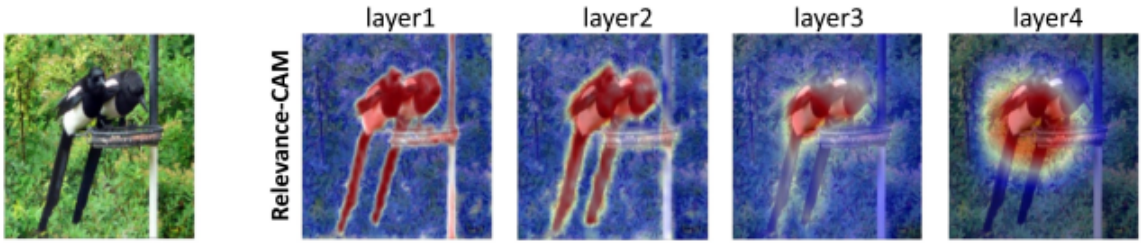

Figure 7. Class sensitivity test of Relevance-CAM on ResNet 50 for single object. The label of the input is a magpie.

클래스별 마스크 이미지 시각화

- 실험 방법: 4.3 섹션에서 설명한 동일한 마스크를 사용하여 '얼룩말'과 '코끼리' 클래스에 대한 마스크된 이미지를 다양한 레이어 깊이에 따라 보여줌
- 결과
  - 마스크된 이미지는 놀라운 클래스 민감도를 보여줌
  - 레이어가 얕아질수록 위치 결정 성능이 약간 저하되지만, 여전히 충분히 클래스 민감한 고해상도 히트맵을 생성함
  - 모델이 이미 1레이어에서 코끼리와 얼룩말을 분리한다는 점에 주목할 수 있음

추가 현상 관찰

- 관찰된 현상
  - 3레이어와 4레이어에서는 오직 까치만이 지역화되며, 2레이어에서는 까치 옆의 막대기가 지역화되기 시작함
  - 1레이어에서는 까치와 까치가 앉아 있는 막대기가 매우 명확하게 함께 지역화됨
- 해석
  - 낮은 레이어에서 막대기를 함께 인식하는 것은, 이미지넷 데이터셋에서 까치가 자주 앉아 있는 것으로 관찰되는 가지와 유사한 막대기를 지역화하는 것으로 이해할 수 있음
  - 즉 얕은 레이어에서 까치와 막대기는 같은 클래스로 간주되며, 레이어가 깊어짐에 따라 까치와 막대기는 별개의 특징으로 분리됨

Figure 6, 7에 대한 요약

- Relevence-CAM의 특징
  - Relevance-CAM은 ResNet 50의 1레이어에서 클래스 특정 히트맵을 생성함
  - 이를 통해 1레이어에서만 일반적인 특징이나 지역 특징이 추출되는 것이 아니라 클래스 특정 정보도 추출된다는 것을 알 수 있음
- 레이어 깊이에 따른 변화
  - 레이어가 깊어짐에 따라 활성화 맵의 채널이 증가함
  - 이에 따라 특징이 얕은 레이어에서 초기에 추출된 범위 내에서 더 세분화됨

4.5 Sanity check for Relevance-CAM

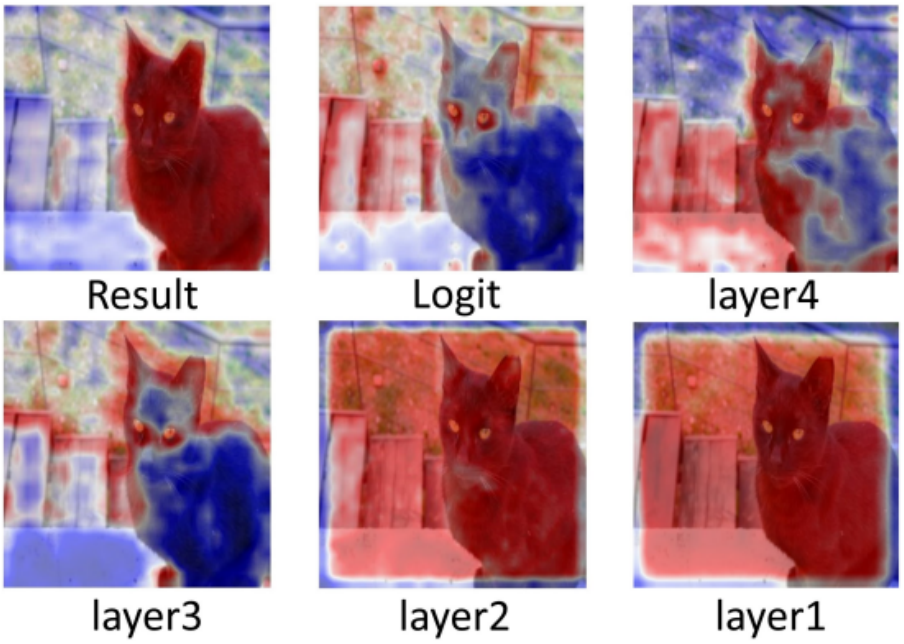

Figure 8. Sanity Check for Relevance-CAM of layer 2 of ResNet50 model

Sanity Check

- 목적: Relevance-CAM 방법의 유효성을 검증하기 위해, 모델 파라미터에 대한 민감도를 평가하는 실험을 수행
- 실험 방법
  - Cascading Randomization Test: 모델의 출력(logit)부터 첫 번째 레이어까지 파라미터를 점진적으로 무작위화(randomizing)하면서 Relevance-CAM 결과를 관찰함
  - 대상 모델: ResNet50 모델의 두 번째 레이어에 대한 Relevance-CAM 결과를 분석함
- 실험 결과
  - 파라미터를 무작위화할수록, 즉 모델의 파라미터가 원래의 값에서 멀어질수록, 주의 맵(attention map)이나 살리언시 맵(saliency map)이 파괴됨
  - Relevance-CAM 방법이 모델 파라미터에 민감하게 반응한다는 것을 의미
- 중요성
  - Relevance-CAM이 모델의 파라미터 변화에 따라 어떻게 반응하는지를 보여줌으로써, 해당 방법이 모델의 결정 과정을 설명하는 데 유효한 도구임을 입증함

- 모델이 어떤 정보에 주목하여 결정을 내리는지를 정확하게 파악할 수 있도록 함

5. Evidence that class specific information is extracted from shallow layer

주요 논점

- 의문점: 얕은 레이어가 주로 일반적인 특징(모서리, 질감 등)을 추출한다면, 특정 클래스의 객체만을 지역화하는 특징 맵이 존재하지 않을 것이라는 의문이 제기될 수 있음
- Relevance-CAM의 접근: Relevance-CAM은 채널별 가중치에 대해서만 LRP(Layer-wise Relevance Propagation)를 사용하기 때문에, 공간별 가중치에는 영향을 미치지 않음

실험 결과

- 얕은 레이어의 결과
  - Relevance-CAM을 사용하여 ResNet50 모델의 얕은 레이어에서도 다른 클래스의 객체들이 분리되어 지역화되는 것을 확인할 수 있음
  - 얕은 레이어에서도 클래스에 특정한 특징 맵이 존재하며, 이러한 특징 맵이 클래스 출력 점수에 큰 영향을 미친다는 것을 시사함
- 깊은 레이어의 결과
  - 깊은 레이어에서의 Relevance-CAM 결과는 얕은 레이어보다 더 명확하게 클래스 특정 특징을 추출할 수 있음을 보여줌

➡️ **얕은 레이어에서도 클래스에 특정한 정보를 추출할 수 있지만, 레이어가 깊어질수록 더 높은 수준의 특징이 추출됨**

➡️ **Relevance-CAM이 모델의 다양한 레이어에서 클래스에 특정한 정보를 효과적으로 추출할 수 있음을 보여줌**

6. Conclusion

Relevance-CAM의 제안

- 목적: 딥러닝 모델과 그 레이어들을 신뢰할 수 있고 정확하게 설명하기 위한 새로운 방법으로 Relevance-CAM을 제안함
- 장점: 다른 설명 방법들이 가진 문제점들(예: shattered gradient 문제, False Confidence)을 보완할 수 있음

Relevance-CAM의 활용

- 얕은 레이어 분석
  - Relevance-CAM을 통해 얕은 레이어에서도 클래스 특정 특징을 추출할 수 있음을 발견함
  - 작은 수용 필드를 가진 얕은 레이어에서도 가능하도록 함
- 다양한 분야의 적용
  - 전이 학습, 모델 프루닝, 약한 지도 학습 분할 등 다양한 분야에 활용될 수 있음
  - 전이 학습 예시: 경험적으로 레이어를 선택하는 대신, 레이어별 분석을 통해 레이어를 선택할 수 있도록 함

➡️ **Relevance-CAM은 다른 연구자들이 딥러닝 모델의 분석을 깊이 있게 할 수 있도록 도우며, 이는 모델의 다양한 레이어에서 중요한 정보를 추출하고 이해하는 데 큰 도움이 될 것으로 보임**

논문에 대한 의견 및 의문점(꼭지)

➡️ Relevance-CAM의 접근 방식이 딥러닝 모델의 해석 가능성을 높이는 데 기여함에도 불구하고, 실제 응용 분야에서의 구체적인 성공 사례나 적용 예시가 더 상세히 제시되었다면 (특정 의료 이미징 분석, 자율 주행 차량의 시각 시스템, 또는 소셜 미디어에서의 이미지 분류 등) 해당 기술의 실용성과 범용성을 더 잘 이해할 수 있었을 것 같아 조금 아쉬운 부분이 있었음.