(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년09월22일
(11) 등록번호 10-2446792
(24) 등록일자 2022년09월20일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) *G06N 3/04* (2006.01)
(52) CPC특허분류
G06N 3/082 (2013.01)
G06N 3/04 (2013.01)
(21) 출원번호 10-2021-0037408
(22) 출원일자 2021년03월23일
심사청구일자 2021년03월23일
(56) 선행기술조사문헌
KR1020170134508 A

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
황도식
서울특별시 서대문구 연세로 50, 제3공학관 C618호(신촌동,연세대학교)
이정룡
서울특별시 서대문구 연세로 50, 제3공학관 C516호(신촌동,연세대학교)
(뒷면에 계속)
(74) 대리인
민영준

전체 청구항 수 : 총 16 항 심사관 : 양대경

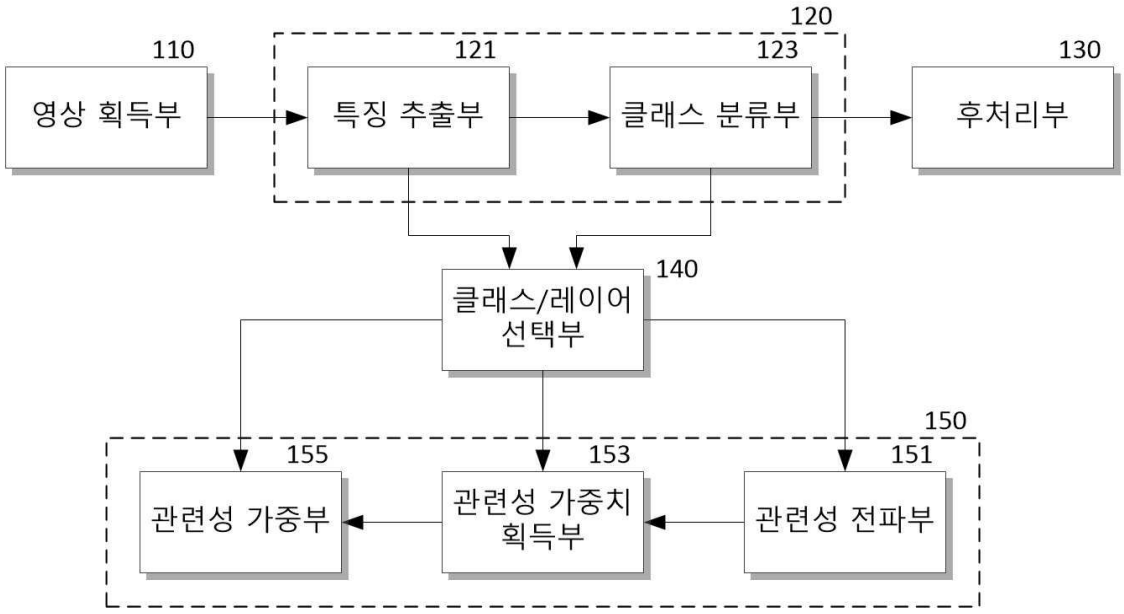
(54) 발명의 명칭 **인공 신경망을 설명하기 위한 관련성 가중 클래스 활성화 맵 생성 장치 및 방법**

(57) 요 약

본 발명은 다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 신경망, 다수의 클래스 중 하

(뒷면에 계속)

대 표 도 - 도1



나의 클래스를 타겟 클래스로 선택하고, 다수의 레이어 중 하나의 레이어를 타겟 레이어로 선택하는 클래스/레이어 선택부 및 신경망의 최종 레이어에서 출력되는 다수의 개별 클래스 분류 결과 중 타겟 클래스에 대응하는 개별 클래스 분류 결과를 초기 관련성으로 설정하고, 설정된 초기 관련성을 최종 레이어로부터 타겟 레이어까지 각 레이어의 가중치 및 활성화 함수에 따라 인접한 이전 레이어로 역전파하여 각 레이어에 대한 관련성 가중치를 획득하며, 획득된 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 관련성 가중 클래스 활성화 맵을 획득하는 관련성 가중 CAM 획득부를 포함하여, 인공 신경망의 특정 레이어가 각 클래스에 대해 주의한 영역을 정확하게 추적하여 관련성 가중 클래스 활성화 맵을 생성하는 클래스 활성화 맵 생성 장치 및 방법을 제공할 수 있다.

(72) 발명자

김세원

서울특별시 서대문구 연세로 50, 제3공학관 C516호(신촌동,연세대학교)

박인용

서울특별시 서대문구 연세로 50, 제3공학관 C516호(신촌동,연세대학교)

어태준

서울특별시 서대문구 연세로 50, 제3공학관 C516호(신촌동,연세대학교)

명 세 서

청구범위

청구항 1

다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 신경망;

상기 다수의 클래스 중 하나의 클래스를 타겟 클래스로 선택하고, 상기 다수의 레이어 중 하나의 레이어를 타겟 레이어로 선택하는 클래스/레이어 선택부; 및

상기 신경망의 최종 레이어에서 출력되는 상기 다수의 개별 클래스 분류 결과 중 상기 타겟 클래스에 대응하는 개별 클래스 분류 결과를 초기 관련성으로 설정하고, 설정된 초기 관련성을 상기 최종 레이어로부터 상기 타겟 레이어까지 각 레이어의 가중치 및 활성화 함수에 따라 인접한 이전 레이어로 역전파하여 각 레이어에 대한 관련성 가중치를 획득하며, 획득된 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 관련성 가중 클래스 활성화 맵을 획득하는 관련성 가중 CAM 획득부를 포함하는 클래스 활성화 맵 생성 장치.

청구항 2

제1항에 있어서, 상기 관련성 가중 CAM 획득부는

상기 초기 관련성을 설정하고, 상기 초기 관련성을 레이어별 관련성 전파(Layer-wise Relevance Propagation: 이하 LRP) 기법에 따라 순차적으로 이전 레이어로 전파하여, 각 레이어에 대한 관련성 맵을 획득하는 관련성 전파부;

각 레이어에 대한 상기 관련성 맵 각각을 글로벌 평균 풀링하여 관련성 가중치를 획득하는 관련성 가중치 획득부; 및

각 레이어에 대한 상기 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 상기 관련성 가중 클래스 활성화 맵을 획득하는 관련성 가중부를 포함하는 클래스 활성화 맵 생성 장치.

청구항 3

제2항에 있어서, 상기 관련성 전파부는

다수의 레이어 중 제l+1 레이어의 각 노드(q)에서 인접한 제l 레이어의 노드(p)로 전파되는 관련성(R_p^l)을 수학식

$$R_p^l = \frac{a_p^l \cdot W_{pq}^{(l,l+1)}}{\sum_m a_m^l \cdot W_{mq}^{(l,l+1)}} R_q^{l+1}$$

(여기서 a_p^l 은 신경망의 제l 레이어의 제p 노드의 활성화값을 나타내며, $W_{pq}^{(l,l+1)}$ 는 제l 레이어의 제p 노드와 제l+1 레이어의 제q 노드 사이의 가중치를 나타내며, m는 제l 레이어의 전체 노드에 대한 인덱스를 나타낸다.)

에 따라 계산하여 전파하는 클래스 활성화 맵 생성 장치.

청구항 4

제3항에 있어서, 상기 관련성 가중치 획득부는

타겟 클래스(c)에 대한 레이어별 관련성 맵(R_k^c)으로부터 각 레이어에 대한 상기 관련성 가중치(α_k^c)를 수학식

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

(여기서 k 는 레이어 식별자를 나타내고, x, y는 제k 레이어에 대한 관련성 맵(R_k^c)의 노드 좌표를 나타낸다.)

에 따라 획득하는 클래스 활성화 맵 생성 장치.

청구항 5

제4항에 있어서, 상기 관련성 가중부는

각 레이어에 대한 상기 관련성 가중치를 대응하는 각 레이어에서 출력된 특징맵과 원소 곱하고, 채널 방향에서 전역 평균 풀링하여 2차원으로 변환하여 각 레이어에 대한 관련성 가중 특징맵을 획득하며, 획득된 각 레이어에 대한 관련성 가중 특징맵을 누적 합산하여 상기 관련성 가중 클래스 활성화 맵을 획득하는 클래스 활성화 맵 생성 장치.

청구항 6

제5항에 있어서, 상기 관련성 가중부는

각 레이어에 대한 관련성 가중 특징맵을 동일한 크기로 변환한 후 누적 합산하는 클래스 활성화 맵 생성 장치.

청구항 7

제4항에 있어서, 상기 관련성 가중부는

상기 관련성 가중 클래스 활성화 맵의 크기를 획득된 영상의 크기로 업스케일링하는 클래스 활성화 맵 생성 장치.

청구항 8

제1항에 있어서, 상기 신경망은

각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하는 다수의 레이어를 포함하는 특징 추출부; 및

상기 특징 추출부에서 최종 출력된 특징맵을 인가받고, 미리 학습된 방식에 따라 최종 출력된 특징맵이 기지정된 다수의 클래스 각각에 대응하는 수준을 계산하여 상기 다수의 개별 클래스 분류 결과를 획득하는 클래스 분류부를 포함하는 클래스 활성화 맵 생성 장치.

청구항 9

다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 신경망을 설명하기 위한 클래스 활성화 맵을 생성하는 클래스 활성화 맵 생성 방법에 있어서,

상기 다수의 클래스 중 하나의 클래스를 타겟 클래스로 선택하고, 상기 다수의 레이어 중 하나의 레이어를 타겟 레이어로 선택하는 단계;

상기 신경망에 영상을 입력하여 다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 단계; 및

상기 신경망의 최종 레이어에서 출력되는 상기 다수의 개별 클래스 분류 결과 중 상기 타겟 클래스에 대응하는 개별 클래스 분류 결과를 초기 관련성으로 설정하고, 설정된 초기 관련성을 상기 최종 레이어로부터 상기 타겟

레이어까지 각 레이어의 가중치 및 활성화 함수에 따라 인접한 이전 레이어로 역전파하여 각 레이어에 대한 관련성 가중치를 획득하며, 획득된 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 관련성 가중 클래스 활성화 맵을 획득하는 단계를 포함하는 클래스 활성화 맵 생성 방법.

청구항 10

제9항에 있어서, 상기 관련성 가중 클래스 활성화 맵을 획득하는 단계는

상기 초기 관련성을 설정하고, 상기 초기 관련성을 레이어별 관련성 전파(Layer-wise Relevance Propagation: 이하 LRP) 기법에 따라 순차적으로 이전 레이어로 전파하여, 각 레이어에 대한 관련성 맵을 획득하는 단계;

각 레이어에 대한 상기 관련성 맵 각각을 글로벌 평균 풀링하여 관련성 가중치를 획득하는 단계; 및

상기 관련성 가중 클래스 활성화 맵을 획득하기 위해 각 레이어에 대한 상기 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하는 단계를 포함하는 클래스 활성화 맵 생성 방법.

청구항 11

제10항에 있어서, 상기 관련성 맵을 획득하는 단계는

다수의 레이어 중 제l+1 레이어의 각 노드(q)에서 인접한 제l 레이어의 노드(p)로 전파되는 관련성(R_p^l)을 수학식

$$R_p^l = \frac{a_p^l \cdot W_{pq}^{(l,l+1)}}{\sum_m a_m^l \cdot W_{mq}^{(l,l+1)}} R_q^{l+1}$$

(여기서 a_p^l 은 신경망의 제l 레이어의 제p 노드의 활성화값을 나타내며, $W_{pq}^{(l,l+1)}$ 는 제l 레이어의 제p 노드와 제l+1 레이어의 제q 노드 사이의 가중치를 나타내며, m는 제l 레이어의 전체 노드에 대한 인덱스를 나타낸다.)

에 따라 계산하여 전파하는 클래스 활성화 맵 생성 방법.

청구항 12

제11항에 있어서, 상기 관련성 가중치를 획득하는 단계는

타겟 클래스(c)에 대한 레이어별 관련성 맵(R_k^c)으로부터 각 레이어에 대한 상기 관련성 가중치(α_k^c)를 수학식

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

(여기서 k 는 레이어 식별자를 나타내고, x, y는 제k 레이어에 대한 관련성 맵(R_k^c)의 노드 좌표를 나타낸다.)

에 따라 획득하는 클래스 활성화 맵 생성 방법.

청구항 13

제12항에 있어서, 상기 특징맵에 가중합하는 단계는

각 레이어에 대한 상기 관련성 가중치를 대응하는 각 레이어에서 출력된 특징맵과 원소곱하고, 채널 방향에서 전역 평균 풀링하여 2차원으로 변환하여 각 레이어에 대한 관련성 가중 특징맵을 획득하는 단계; 및

상기 관련성 가중 클래스 활성화 맵을 각 레이어에 대한 관련성 가중 특징맵을 누적 합산하여 획득하는 단계를 포함하는 클래스 활성화 맵 생성 방법.

청구항 14

제13항에 있어서, 상기 특징맵에 가중합하는 단계는

상기 누적 합산하여 획득하는 단계 이전, 각 레이어에 대한 관련성 가중 특징맵을 동일한 크기로 변환하는 단계를 더 포함하는 클래스 활성화 맵 생성 방법.

청구항 15

제14항에 있어서, 상기 특징맵에 가중합하는 단계는

상기 누적 합산하여 획득하는 단계 이후, 상기 관련성 가중 클래스 활성화 맵의 크기를 획득된 영상의 크기로 업스케일링하는 단계를 더 포함하는 클래스 활성화 맵 생성 방법.

청구항 16

제9항에 있어서, 상기 다수의 개별 클래스 분류 결과를 획득하는 단계는

각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 각 레이어에 대응하는 다수의 특징맵을 출력하는 단계; 및

최종 출력된 특징맵을 인가받고, 미리 학습된 방식에 따라 최종 출력된 특징맵이 기지정된 다수의 클래스 각각에 대응하는 수준을 계산하여 상기 다수의 개별 클래스 분류 결과를 획득하는 단계를 포함하는 클래스 활성화 맵 생성 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 클래스 활성화 맵 생성 장치 및 방법에 관한 것으로, 인공 신경망을 설명하기 위한 관련성 가중 클래스 활성화 맵 생성 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 인공 신경망에 대한 응용 분야가 증가하고, 이에 다양한 인공 신경망이 개발됨에 따라 인공 신경망(또는 딥 러닝 모델)을 설명하는 능력도 점점 더 중요 해지고 있다. 특히 인공 신경망의 실제 적용에 앞서 딥 러닝 모델의 추론과 결과 생성 과정을 분석하는 것이 중요하다.

[0003] 일 예로 의료 분야에 이용되는 인공 신경망으로 의료 영상으로부터 암 발생 여부를 진단하는 인공 신경망의 경우, 해당 인공 신경망이 어떠한 추론 과정과 결과 생성 과정을 통해 암 진단을 수행하였는지를 분석할 필요가 있다. 이는 학습 완료된 것으로 판단되어 암 진단을 수행하는 인공 신경망이 실제로는 암 병변 부위가 아닌 다른 부위를 기준으로 암 발생 여부를 판별하여 오진을 수행했을 가능성이 있기 때문이다. 즉 암 진단을 수행한 인공 신경망이 실제로 암 병변 영역에 주의를 기울여 암을 진단하였는지 여부를 검증할 수 있도록 인공 신경망의 동작을 분석할 필요가 있다.

[0004] 이와 같이 딥 러닝 모델의 추론과 결과 생성 과정을 분석하는 학문을 설명가능 인공 지능(Explainable AI: 이하 XAI)라 하며, 이러한 XAI에서 대표적인 방법으로 클래스 활성화맵(Class Activation Mapping: CAM)을 이용하는 방법이 있다. CAM 기반 방법은 컴퓨터 비전(Computer Vision) 분야에서 널리 사용되는 컨볼루션 신경망(Convolutional Neural Network: 이하 CNN)을 설명하기 위해 이용되는 방법으로, CNN에 의해 구분되는 각 클래스를 분류하기 위해 CNN이 주목한 영역을 시각적으로 표시하는 기법이다. 이때, CAM 기반 방법은 적어도 하나의 컨볼루션 레이어를 포함하는 CNN의 마지막 레이어가 분류된 클래스값을 획득하기 위해 주목한 영역을 추적한다.

[0005] 다만 CAM 기반 방법은 CNN에서 마지막 레이어가 주목한 영역만을 추적하므로, 하나의 컨볼루션 레이어만으로 구성된 CNN에서는 효과적이나, 다수의 컨볼루션 레이어를 포함하는 CNN에 적용하기 어렵다. 이에 현재는 주로 Grad-CAM 또는 Gram-CAM++ 이 이용되고 있다. CAM이 특정 컨볼루션 레이어에서 추출된 특징맵(feature map)이므로, Grad-CAM 또는 Gram-CAM++는 CAM과 특정 레이어에에서 획득된 특징맵의 각 픽셀 사이 변화를 추정하여 기울기 맵(Gradient map)을 획득하고, 획득된 기울기 맵을 이용하여 특징맵에 대한 가중치를 계산하여 가중함으로써, 가중 클래스 활성화 맵(gradient-weighted Class Activation Map)을 생성하는 방식이다.

[0006] 그러나 Grad-CAM 또는 Gram-CAM++ 기법은 CAM과 선택된 특정 특징맵 사이의 기울기를 곧바로 추정하여 기울기

맵을 획득하므로, CAM과 선택된 레이어 사이의 레이어의 개수가 증가할수록, CAM과 선택된 특정 특징맵을 출력하는 레이어 사이에 배치되는 다수의 레이어의 동작 특성을 제대로 반영할 수 없어 추정된 기울기에 많은 노이즈가 포함되고 불연속적이 되어 그 성능이 크게 저하되는 문제가 있다. 이를 부서진 그래디언트 문제(shattered gradient problem)라 하며, 이와 같은 부서진 그래디언트 문제로 인해 Grad-CAM 또는 Gram-CAM++를 대체하여 다수의 레이어가 포함된 인공 신경망의 다수의 레이어 각각이 각 클래스에 대해 주목한 영역을 정확하게 분석할 수 있는 새로운 기법이 요구되고 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 한국 공개 특허10-2021-0002018호 (2021.01.06 공개)

발명의 내용

해결하려는 과제

[0008] 본 발명의 목적은 인공 신경망의 다수의 레이어 각각이 각 클래스에 대해 주의한 영역을 정확하게 추정할 수 있는 클래스 활성화 맵 생성 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0009] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른 클래스 활성화 맵 생성 장치는 다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 신경망; 상기 다수의 클래스 중 하나의 클래스를 타겟 클래스로 선택하고, 상기 다수의 레이어 중 하나의 레이어를 타겟 레이어로 선택하는 클래스/레이어 선택부; 및 상기 신경망의 최종 레이어에서 출력되는 상기 다수의 개별 클래스 분류 결과 중 상기 타겟 클래스에 대응하는 개별 클래스 분류 결과를 초기 관련성으로 설정하고, 설정된 초기 관련성을 상기 최종 레이어로부터 상기 타겟 레이어까지 각 레이어의 가중치 및 활성화 함수에 따라 인접한 이전 레이어로 역전파하여 각 레이어에 대한 관련성 가중치를 획득하며, 획득된 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 관련성 가중 클래스 활성화 맵을 획득하는 관련성 가중 CAM 획득부를 포함한다.

[0010] 상기 관련성 가중 CAM 획득부는 상기 초기 관련성을 설정하고, 상기 초기 관련성을 레이어별 관련성 전파(Layer-wise Relevance Propagation: 이하 LRP) 기법에 따라 순차적으로 이전 레이어로 전파하여, 각 레이어에 대한 관련성 맵을 획득하는 관련성 전파부; 각 레이어에 대한 상기 관련성 맵 각각을 글로벌 평균 풀링하여 관련성 가중치를 획득하는 관련성 가중치 획득부; 및 각 레이어에 대한 상기 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 상기 관련성 가중 클래스 활성화 맵을 획득하는 관련성 가중부를 포함할 수 있다.

[0011] 상기 관련성 전파부는 다수의 레이어 중 제l+1 레이어의 각 노드(q)에서 인접한 제l 레이어의 노드(p)로 전파되는 관련성(R_p^l)을 수학식

$$R_p^l = \frac{a_p^l \cdot W_{pq}^{(l,l+1)}}{\sum_m a_m^l \cdot W_{mq}^{(l,l+1)}} R_q^{l+1}$$

[0012]

[0013] (여기서 a_p^l 은 신경망의 제l 레이어의 제p 노드의 활성화값을 나타내며, $W_{pq}^{(l,l+1)}$ 는 제l 레이어의 제p 노드와 제l+1 레이어의 제q 노드 사이의 가중치를 나타내며, m는 제l 레이어의 전체 노드에 대한 인덱스를 나타낸다.)에 따라 계산하여 전파할 수 있다.

[0014] 상기 관련성 가중치 획득부는 타겟 클래스(c)에 대한 레이어별 관련성 맵(R_k^c)으로부터 각 레이어에 대한 상기

관련성 가중치(α_k^c)를 수학식

[0015] $$\alpha_k^c = \sum_{x,y} R_k^c(x,y)$$

[0016] (여기서 k 는 레이어 식별자를 나타내고, x, y는 제k 레이어에 대한 관련성 맵(R_k^c)의 노드 좌표를 나타낸다.)에 따라 획득할 수 있다.

[0017] 상기 관련성 가중부는 각 레이어에 대한 상기 관련성 가중치를 대응하는 각 레이어에서 출력된 특징맵과 원소 곱하고, 채널 방향에서 전역 평균 풀링하여 2차원으로 변환하여 각 레이어에 대한 관련성 가중 특징맵을 획득하며, 획득된 각 레이어에 대한 관련성 가중 특징맵을 누적 합산하여 상기 관련성 가중 클래스 활성화 맵을 획득할 수 있다.

[0018] 상기 관련성 가중부는 각 레이어에 대한 관련성 가중 특징맵을 동일한 크기로 변환한 후 누적 합산할 수 있다.

[0019] 상기 관련성 가중부는 상기 관련성 가중 클래스 활성화 맵의 크기를 획득된 영상의 크기로 업스케일링할 수 있다.

[0020] 상기 목적을 달성하기 위한 본 발명의 다른 실시예에 따른 클래스 활성화 맵 생성 방법은 다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 신경망을 설명하기 위한 클래스 활성화 맵을 생성하는 클래스 활성화 맵 생성 방법에 있어서,

[0021] 상기 다수의 클래스 중 하나의 클래스를 타겟 클래스로 선택하고, 상기 다수의 레이어 중 하나의 레이어를 타겟 레이어로 선택하는 단계; 상기 신경망에 영상을 입력하여 다수의 레이어를 포함하는 각각 입력되는 영상 또는 이전 레이어에서 출력된 특징맵을 인가받아 미리 학습된 방식에 따라 특징을 추출하여 특징맵을 출력하고, 최종 출력된 특징맵을 미리 학습된 방식에 따라 분류하여 기지정된 다수의 클래스에 대응하는 다수의 개별 클래스 분류 결과를 획득하는 단계; 및 상기 신경망의 최종 레이어에서 출력되는 상기 다수의 개별 클래스 분류 결과 중 상기 타겟 클래스에 대응하는 개별 클래스 분류 결과를 초기 관련성으로 설정하고, 설정된 초기 관련성을 상기 최종 레이어로부터 상기 타겟 레이어까지 각 레이어의 가중치 및 활성화 함수에 따라 인접한 이전 레이어로 역전파하여 각 레이어에 대한 관련성 가중치를 획득하며, 획득된 관련성 가중치를 대응하는 레이어에서 출력된 특징맵에 가중합하여 관련성 가중 클래스 활성화 맵을 획득하는 단계를 포함한다.

발명의 효과

[0022] 따라서, 본 발명의 실시예에 따른 클래스 활성화 맵 생성 장치 및 방법은 클래스 활성화 맵을 레이어별 전파 기법에 따라 각 레이어의 배치 순서에 역순에 따라 순차적으로 전파하여 인공 신경망의 특정 레이어가 각 클래스에 대해 주의한 영역을 정확하게 추적하여 관련성 가중 클래스 활성화 맵을 생성할 수 있도록 한다. 그러므로 깊은 깊이를 갖는 인공 신경망의 클래스 분류에 대한 추론 및 결과 생성 과정을 정확하게 설명하여 분석할 수 있도록 한다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 일 실시예에 따른 클래스 활성화 맵 생성 장치의 개략적 구조를 나타낸다.

도 2는 도 1의 클래스 활성화 맵 생성 장치의 동작을 설명하기 위한 도면이다.

도 3은 인공 신경망의 일 예에 대한 동작을 설명하기 위한 도면이다.

도 4 및 도 5는 도 3의 인공 신경망에서 컨볼루션 커널의 동작을 설명하기 위한 도면이다.

도 6은 관련성 전파부의 동작을 설명하기 위한 도면이다.

도 7 내지 도 8은 본 실시예에 따른 클래스 활성화 맵 생성 장치의 성능을 나타낸다.

도 9는 본 발명의 일 실시예에 따른 클래스 활성화 맵 생성 방법을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 발명과 본 발명의 동작상의 이점 및 본 발명의 실시에 의하여 달성되는 목적을 충분히 이해하기 위해서는 본 발명의 바람직한 실시예를 예시하는 첨부 도면 및 첨부 도면에 기재된 내용을 참조하여야만 한다.

[0025] 이하, 첨부한 도면을 참조하여 본 발명의 바람직한 실시예를 설명함으로써, 본 발명을 상세히 설명한다. 그러나, 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며, 설명하는 실시예에 한정되는 것이 아니다. 그리고, 본 발명을 명확하게 설명하기 위하여 설명과 관계없는 부분은 생략되며, 도면의 동일한 참조부호는 동일한 부재임을 나타낸다.

[0026] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라, 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 "...부", "...기", "모듈", "블록" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0027] 도 1은 본 발명의 일 실시예에 따른 클래스 활성화 맵 생성 장치의 개략적 구조를 나타내고, 도 2는 도 1의 클래스 활성화 맵 생성 장치의 동작을 설명하기 위한 도면이다.

[0028] 도 1을 참조하면, 본 실시예에 따른 클래스 활성화 맵 생성 장치는 영상 획득부(110), 신경망(120), 후처리부(130), 클래스/레이어 선택부(140) 및 관련성 가중 CAM 획득부(150)를 포함할 수 있다.

[0029] 영상 획득부(110)는 사전에 학습된 신경망(120)의 추론과 결과 생성 과정, 즉 동작을 분석하기 위한 영상을 획득한다. 영상 획득부(110)는 카메라 또는 CCTV와 같이 영상을 획득할 수 있는 영상 획득 장치로 구현되거나, 미리 획득된 영상이 저장된 메모리 등으로 구현될 수 있다. 또한 경우에 따라서는 다른 장치로부터 네트워크를 통해 영상을 전송받는 통신 장치 등으로 구현될 수도 있다.

[0030] 신경망(120)은 딥 러닝 모델로 구현되어 미리 학습된 인공 신경망으로서, 영상 획득부(110)에서 획득한 영상을 인가받아 미리 학습된 방식에 따라 인가된 영상의 특징을 추출하여 특징맵을 획득하고, 획득된 특징맵의 클래스를 분류한다. 이때, 신경망(120)은 영상이 기지정된 다수의 클래스 각각에 대응하는 수준을 추출하는 방식으로 분류할 수 있다.

[0031] 신경망(120)은 특징 추출부(121) 및 클래스 분류부(123)를 포함할 수 있다. 특징 추출부(121)는 미리 학습된 방식에 따라 인가되는 영상의 특징을 추출하여 특징맵을 획득한다. 여기서 특징 추출부(121)는 학습에 의해 가중치가 설정된 다수의 레이어를 포함하고, 다수의 레이어 각각은 영상 또는 이전 레이어에서 획득한 특징맵을 인가받아 설정된 가중치를 적용하여 대응하는 특징맵을 출력하도록 구성된다.

[0032] 클래스 분류부(123)는 특징 추출부(121)에서 최종 획득된 최종 특징맵을 인가받고, 미리 학습된 방식에 따라 최종 특징맵이 기지정된 다수의 클래스 각각에 대응하는 수준을 계산하여 클래스 분류 결과(y)를 도출한다. 여기서 클래스 분류 결과(y)는 기지정된 다수의 클래스 각각에 대응 수준을 나타내는 개별 클래스 분류 결과(y^c)의 집합으로 획득될 수 있으며, 개별 클래스 분류 결과(y^c) 각각은 수치값으로 획득될 수 있다.

[0033] 한편 후처리부(130)는 신경망(120)에서 획득된 개별 클래스 분류 결과(y^c)를 기반으로 신경망(120)의 활용 목적에 따른 미리 지정된 동작을 수행한다. 일 예로 후처리부(130)는 개별 클래스 분류 결과(y^c)에 기반하여 각 클래스에 대응하는 객체가 영상에 포함되어 있는지 여부를 식별하거나, 암과 같은 특정 질병의 발생 여부를 판단할 수 있다. 후처리부(130)는 일 예로 개별 클래스 분류 결과(y^c)를 기지정된 방식에 따라 확률값으로 변환하여 객체의 포함 여부를 식별하도록 구성될 수 있다.

[0034] 본 실시예에서 신경망(120)은 도 2에 도시된 바와 같이, 일 예로 CNN 또으로 구현될 수 있으며, CNN으로부터 파생된 ResNet(Residual Network) 등으로 구현될 수 있다. 특징 추출부(121)는 n개 컨볼루션 레이어를 포함하여 n개의 특징맵(A_1^C, A_2^C, …, A_n^C)을 획득할 수 있다. 그리고 클래스 분류부(123) 또한 적어도 하나의 레이어를 포함하여 구성될 수 있으며, 클래스 분류부(123)는 신경망(120)의 활용 목적에 따라 개별 클래스 분류 결과(y^c)를 추출하기 위한 다양한 구조의 네트워크로 구성될 수 있다. 여기서 클래스 분류부(123)의 적어도 하나의 각각은 일 예로 FC 레이어(Fully Connected layer)로 구성될 수 있다.

[0035] 도 1에서는 설명의 편의를 위하여 후처리부(130)를 신경망(120)과 별도의 구성으로 도시하였으나, 신경망(120)은 후처리부(130)를 포함하여 구성될 수도 있다.

[0036] 클래스/레이어 선택부(140)는 신경망(120)이 분류할 수 있는 기지정된 다수의 클래스 중 하나의 클래스를 타겟 클래스(c)로 선택하고, 특징 추출부(121)의 다수의 레이어 중 하나의 레이어를 타겟 레이어(t)로 선택한다. 여기서 클래스/레이어 선택부(140)는 사용자 명령에 따라 타겟 클래스(c) 및 타겟 레이어(t)를 선택할 수 있다.

[0037] 관련성 가중 CAM 획득부(150)는 특징 추출부(121)의 다수의 레이어 중 클래스/레이어 선택부(140)에 의해 선택된 타겟 레이어(t)가 다수의 클래스 중 타겟 클래스(c)를 분류하기 위해 입력된 영상 또는 특징맵에서 어느 부분에 대해 주의하였는지를 나타내는 CAM을 획득한다. 이때 관련성 가중 CAM 획득부(150)는 기존의 Grad-CAM 또는 Gram-CAM++ 기법과 달리 선택된 클래스에 대응하는 개별 클래스 분류 결과(y^c)와 선택된 레이어에서 출력되는 특징맵 사이의 기울기를 곧바로 계산하지 않는다.

[0038] 대신 본 실시예의 관련성 가중 CAM 획득부(150)는 신경망(120)의 다수의 레이어 사이의 관련성(Relevance)(R)을 최종 레이어로부터 선택된 타겟 레이어(t)까지 순차적으로 전파시킨다. 즉 각 레이어의 동작을 역순으로 추적하여 각 레이어 사이의 관련성을 획득한다. 이와 같이 신경망(120)을 다수의 레이어 단위로 분할하고, 분할된 다수의 레이어 중 최종 레이어로부터 순차적으로 인접한 이전 레이어와의 관련성(R)을 순차 전파하는 기법을 여기서는 레이어별 관련성 전파(Layer-wise Relevance Propagation: 이하 LRP) 기법이라 한다.

[0039] Grad-CAM 또는 Gram-CAM++ 기법과 같이 신경망의 출력인 개별 클래스 분류 결과(y^c)와 선택된 레이어 사이의 관련성을 곧바로 추정하는 기법의 경우, 부서진 그래디언트 문제로 인해 레이어의 수에 따라 추정된 레이어까지의 관련성에 대한 오차가 크게 증가하지만, 개별 클래스 분류 결과(y^c)와 선택된 레이어 사이의 관련성(R)을 LRP 기법에 따라 레이어 단위로 역순으로 순차 추정하는 경우, 레이어의 개수가 증가하더라도 추정되는 관련성에 오차가 증가하는 것을 최대한 억제할 수 있다.

[0040] 그리고 관련성 가중 CAM 획득부(150)는 최종 레이어로부터 타겟 레이어(t)까지 전파되어 각 레이어로 획득된 관련성(R)에 기반하여, 타겟 레이어(t)로부터 최종 레이어까지의 각 레이어가 타겟 클래스(c)에 대한 개별 클래스 분류 결과(y^c)를 도출하기 위해 수행한 영역별 기여도를 나타내는 관련성 가중치(α)를 획득한다.

[0041] 상기한 부서진 그래디언트 문제는 각 컨볼루션 레이어에 포함된 활성화 함수(activation function)를 고려하지 않는 것이 주요 오차로 발생하므로, LRP 기법을 적용하는 본 실시예에서는 레이어 단위로 관련성을 역전파하면서 각 레이어의 활성화 함수를 함께 고려함으로써, 오차의 증가를 최소화할 수 있다.

[0042] 그리고 관련성 가중 CAM 획득부(150)는 선택된 타겟 레이어(t)로부터 최종 레이어까지 각 레이어에서 획득된 특징맵(A_k^c)에 대응하는 각 레이어에 대해 획득된 관련성 가중치(α_k)를 가중합하여 선택된 레이어가 영상의 각 영역별로 선택된 클래스를 분류하기 위해 주의한 수준을 나타내는 관련성 가중 클래스 활성화 맵($L_{Relevance-CAM}$)을 획득한다.

[0043] 관련성 가중 CAM 획득부(150)는 관련성 전파부(151), 관련성 가중치 획득부(153) 및 관련성 가중부(155)를 포함할 수 있다.

[0044] 관련성 전파부(151)는 클래스/레이어 선택부(140)에서 선택된 개별 클래스 분류 결과(y^c)와 신경망(120)의 최종 레이어로부터 선택된 타겟 레이어(t)까지의 각 레이어의 가중치 및 활성화 함수를 인가받고, 인가된 개별 클래스 분류 결과(y^c)를 신경망(120)의 각 레이어의 가중치 및 활성화 함수를 기반으로 레이어의 배치 순서의 역순으로 전파하여, 타겟 레이어(t)까지 인접한 레이어의 픽셀들(노드) 사이의 관련성(R_p)을 계산하여 관련성 맵(R)을 획득한다.

[0045] 관련성 가중치 획득부(153)는 관련성 전파부(151)에서 획득된 각 레이어에 대한 관련성 맵(R_k)의 관련성 값(R_p)들을 글로벌 평균 풀링(Global Average Pooling)하여 관련성 가중치(α_k)를 획득한다.

[0046] 그리고 관련성 가중부(155)는 관련성 가중치 획득부(153)에서 타겟 레이어(t)로부터 최종 레이어까지 각 레이어에 대해 획득된 관련성 가중치(α_k)와 신경망(120)의 각 레이어에서 출력된 특징맵(A_k^c)을 원소곱하여 가중하여,

관련성 가중 특징맵을 획득하고, 관련성 가중 특징맵을 누적 합산하여 관련성 가중 클래스 활성화 맵($L_{Relevance-CAM}$)을 획득한다. 여기서 관련성 가중부(155)는 타겟 레이어(t)에서 출력된 특징맵이 3차원인 경우, 3차원의 특징맵 또한 채널 방향으로 글로벌 평균 풀링한 후 관련성 가중치를 원소곱하거나, 3차원의 특징맵에 관련성 가중치를 원소곱한 이후, 채널 방향으로 글로벌 평균 풀링하여 2차원으로 변환하여 관련성 가중 특징맵을 획득하고, 관련성 가중 특징맵을 누적 합산하여 관련성 가중 클래스 활성화 맵을 획득할 수도 있다.

[0047] 따라서 관련성 가중 클래스 활성화 맵은 타겟 레이어가 타겟 클래스의 특징을 추출함에 있어 입력된 영상 또는 특징맵에서 어느 영역에 주의를 기울였는지 나타내는 2차원 맵으로 획득될 수 있다. 다만 신경망(120)의 다수의 레이어 각각이 입력되는 영상 또는 특징맵에서 특징을 추출하여 특징맵을 생성하는 경우, 특징맵의 크기가 입력 영상에 대비하여 작아진다.

[0048] 이에 관련성 가중부(155)는 각 레이어에 대해 획득된 2차원의 관련성 가중 특징맵이 동일한 2차원 크기로 변환한 후 합산할 수 있다. 또한 영상에서 레이어의 관심 영역을 정확히 확인할 수 있도록, 관련성 가중부(155)는 관련성 가중 클래스 활성화 맵의 크기를 영상 획득부(110)에서 획득한 영상의 크기로 업스케일링할 수 있다.

[0049] 이하에서는 관련성 가중 CAM 획득부(150)의 동작을 도면을 참조하여 더욱 상세하게 설명한다. 다만 관련성 가중 CAM 획득부(150)의 동작을 설명하기 위해서는 신경망(120)의 동작을 우선 살펴볼 필요가 있다. 이에 신경망(120)의 동작을 우선 설명하고, 이후, 관련성 가중 CAM 획득부(150)의 동작을 설명한다.

[0050] 도 3은 인공 신경망의 일 예에 대한 동작을 설명하기 위한 도면이고, 도 4 및 도 5는 도 3의 인공 신경망에서 컨볼루션 커널의 동작을 설명하기 위한 도면이다.

[0051] 도 3은 인공 신경망의 일 예로 CNN을 도시하였으며, CNN에서 영상 인식, 음성 인식, 자연어 처리, 필기체 인식 등에 이용되는 LeNET-5를 도시하였다. 도 3에 도시된 바와 같이, LeNET-5는 32 X 32 크기의 입력 영상(Input)을 인가받아, 컨볼루션 연산 및 서브 샘플링 연산(여기서는 풀링 연산)을 반복적으로 수행하며 특징 맵(feature map)을 추출하고, 특징 맵에서 추출된 특징을 기반으로 기지정된 클래스 중 가장 가능성이 큰 클래스에 대응하는 값을 선택하도록 구성된다.

[0052] 여기서 특징 추출 동작(Feature extraction)은 특징 추출부(121)에 의해 수행되며, 컨볼루션 연산과 서브 샘플링 연산은 동일한 컨볼루션 레이어(C1, C2, C3)에서 수행된다. 그리고 분류 동작(Classification) 동작은 클래스 분류부(123)에 의해 수행되며, 2개의 완전 연결 레이어(Fully Connected Layer)(FC1, FC2)에서 수행된다. 즉 도 3의 신경망(120)은 5개의 레이어로 구성되며, 32 X 32 크기의 입력 영상(Input)을 인가받아 기지정된 10개의 클래스 중 적어도 하나의 클래스로 분류하여 출력한다.

[0053] 그리고 다수의 레이어(C1, C2, C3, FC1, FC2) 각각은 학습에 의해 설정된 다수의 가중치를 포함하는 커널(Kernel)을 포함하여 구성된다.

[0054] 도 4에 도시된 바와 같이, 커널을 포함하는 다수의 레이어 각각은 입력 영상 또는 이전 레이어에서 획득된 특징맵을 인가받아 컨볼루션 연산과 같이 기지정된 연산을 수행하여 출력 특징맵을 출력한다.

[0055] 이때 커널은 도 5에 도시된 바와 같이, 입력되는 영상 또는 특징맵의 각 픽셀(또는 원소)(x)을 대응하는 가중치(w)로 가중합($\sum xw$)하고, 기지정된 활성화 함수(ϕ)로 활성화하여 출력되는 특징맵의 각 원소인 특징(o)을 추출한다.

[0056] 도 6은 관련성 전파부의 동작을 설명하기 위한 도면이다.

[0057] 도 6에 도시된 바와 같이, 본 실시예에서 관련성 전파부(151)는 LRP 기법에 기반하여, 관련성을 최종 레이어로부터 선택된 타겟 레이어(t)까지 레이어 단위로 역방향으로 순차적으로 역전파할 수 있다.

[0058] LRP 기법에 따르면 제l+1 레이어의 각 노드(q)에서 인접한 제l 레이어의 노드(p)로 전파되는 관련성(R_p^l)은 수학식 1에 따라 전파될 수 있다.

수학식 1

$$R_p^l = \frac{a_p^l \cdot W_{pq}^{(l,l+1)}}{\sum_m a_m^l \cdot W_{mq}^{(l,l+1)}} R_q^{l+1}$$

[0059]

[0060] 여기서 R_p^l 은 제l 레이어의 제p 노드의 관련성 값(Relevance value), R_q^{l+1} 은 제l+1 레이어의 제q 노드의 관련성 값을 나타내고, a_p^l 은 신경망의 제l 레이어의 제p 노드의 활성화값(activation value)을 나타내며, $W_{pq}^{(l,l+1)}$ 는 제l 레이어의 제p 노드와 제l+1 레이어의 제q 노드 사이의 가중치를 나타낸다. 그리고 m는 제l 레이어의 전체 노드에 대한 인덱스를 나타낸다.

[0061] 수학식 1은 제l+1 레이어의 제q 노드의 값을 제l 레이어의 제p 노드의 값으로 분해할 수 있도록 한다. 수학식 1에서 분자항은 제l 레이어의 제p 노드와 제l+1 레이어의 제q 노드 사이의 가중치로 고려될 수 있으며, 분모항은 제l+1 레이어의 제q 노드와 연결된 제l 레이어의 모든 노드와의 가중치와 활성화 값을 곱한 값의 합으로 제l+1 레이어의 관련성(R_q^{l+1})을 분배하기 위한 정규화 항목(normalize term)으로 볼 수 있다.

[0062] 이에 초기 관련성이 설정되면, 수학식 1에 따라 설정된 초기 관련성을 타겟 레이어(t)까지 이전 레이어로 순차적으로 전파하여, 타겟 레이어(t)에 대한 관련성을 모두 계산함으로써 관련성 맵(R)을 획득할 수 있다. 관련성 전파부(151)는 신경망(120)의 클래스 분류 결과(y) 중 선택된 타겟 클래스(c)에 대응하는 개별 클래스 분류 결과(y^c)를 타겟 클래스에 대한 초기 관련성(R^c)으로 설정할 수 있다.

[0063] 관련성 가중치 획득부(153)는 관련성 전파부(151)가 타겟 클래스(c)에 대해 각 레이어에 대해 획득한 관련성 맵(R_k^c)을 수학식 2에 따라 전역 평균 풀링하여 각 레이어(k)에서 출력되는 특징맵(A_k^c)들 각각에 대한 관련성 가중치(α_k^c)를 계산한다.

수학식 2

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

[0064]

[0065] 여기서 k 는 레이어 식별자를 나타내고, x, y는 제k 레이어에 대한 관련성 맵(R_k^c)의 노드 좌표를 나타낸다.

[0066] 수학식 2에 따라 타겟 클래스(c)에 대한 각 레이어(k)의 관련성 가중치(α_k^c)가 계산되면, 관련성 가중부(155)는 계산된 관련성 가중치(α_k^c)를 대응하는 각 레이어(k)에서 출력된 특징맵(A_k^c)에 가중하여 관련성 가중 특징맵($\alpha_k^c A_k^c$)을 획득하고, 획득된 관련성 가중 특징맵($\alpha_k^c A_k^c$)을 합산하여 타겟 클래스(c) 및 타겟 레이어(k)에 대한 가중 클래스 활성화 맵($L_{Relevance-CAM}$)을 획득한다.

[0067] 관련성 가중부(155)는 수학식 3에 따라 가중치(α_k^c)와 특징맵(A_k^c)으로부터 가중 클래스 활성화 맵($L_{Relevance-CAM}$)을 획득할 수 있다.

수학식 3

[0068] $$L^{c}_{Relevance-CAM} = \sum_{k} \alpha^{c}_{k} A^{c}_{k}$$

[0069] 그리고 관련성 가중부(155)는 수학식 3에 따라 획득된 관련성 가중 클래스 활성화 맵($L_{Relevance-CAM}$)의 크기를 신경망(120)에 입력되는 영상의 크기로 업스케일링하여 출력할 수 있다.

[0070] 도 7 및 도 8은 본 실시예에 따른 클래스 활성화 맵 생성 장치의 성능을 나타낸다.

[0071] 도 7 내지 도 8에서는 본 실시예에 따른 클래스 활성화 맵 생성 장치의 성능을 비교하기 위해, 본 실시예에 따라 획득되는 관련성 가중 클래스 활성화 맵(Relevance-CAM)과 기존의 클래스 활성화 맵 생성 기법인 Grad-CAM, Grad-CAM++ 및 Score-CAM에 의해 획득된 클래스 활성화 맵을 함께 도시하였다.

[0072] 그리고 도 7에서는 신경망을 구성하는 4개의 레이어(layer1 ~ layer4) 각각에 대한 CAM을 도시하였다. 도 7을 살펴보면, 최종 레이어인 제4 레이어(layer4)에 대한 CAM은 모두 유사하게 입력 영상에서 나비가 포함된 영역에 대해 주목하고 있음을 알 수 있다. 그러나 기존의 클래스 활성화 맵 생성 기법인 Grad-CAM, Grad-CAM++ 및 Score-CAM에서는 제2 레이어(layer2)에 대한 CAM에서 이미 나비 영역이 아닌 영역에도 상당부분 주목하고 있을 뿐만 아니라, 나비 영역에서도 일부만을 주목하고 있음을 알 수 있다. 이는 실제 제2 레이어(layer2)가 주목한 영역을 나타내고 있는 것일 수도 있으나, 기존의 클래스 활성화 맵 생성 기법의 오류로 인해 제2 레이어(layer2)가 주목한 영역을 정확하게 추출하지 못해서 발생한 문제일 수도 있다.

[0073] 그에 비해 본 실시예에 따른 클래스 활성화 맵 생성 장치가 획득한 CAM의 경우, 제2 레이어(layer2) 뿐만 아니라 제1 레이어(layer1)에서 이미 나비 영역이 정확하게 도출됨을 알 수 있다. 이로서 제1 레이어(layer1)가 이미 나비 영역을 정확하게 주목하였음에도 기존의 클래스 활성화 맵 생성 기법이 이를 제대로 분석하지 못한 것임을 알 수 있다.

[0074] 한편, 도 8은 다양한 객체가 포함된 영상에 대한 레이어별 CAM을 나타낸다. 도 8에서도 기존의 클래스 활성화 맵 생성 기법인 Grad-CAM, Grad-CAM++ 및 Score-CAM에 의해 획득된 CAM과 비교하여 도시하였다. 도 8을 참조하면, 본 실시예에 따른 클래스 활성화 맵 생성 장치는 다양한 객체가 포함된 영상, 특히 왼쪽 아래와 같이 다수의 객체가 포함된 영상에서도 각 레이어가 주목한 영역을 나타내는 CAM을 정확하게 추출할 수 있다.

[0075] 도 9는 본 발명의 일 실시예에 따른 클래스 활성화 맵 생성 방법을 나타낸다.

[0076] 도 1 내지 도 6을 참조하여, 본 실시예에 따른 클래스 활성화 맵 생성 방법을 설명하면, 우선 신경망에서 분석하고자 하는 타겟 클래스(c) 및 타겟 레이어(t)를 선택한다(S10). 그리고 타겟 클래스(c)가 포함된 영상을 획득한다(S20). 영상이 획득되면, 신경망(120)의 특징 추출부(121)가 획득된 영상을 인가받고, 미리 학습된 방식에 따라 특징을 추출하여 특징맵(A_k^c)을 획득한다(S30). 이때 신경망(120)은 다수의 레이어를 포함하여 구성될 수 있으며, 다수의 레이어 각각은 영상 또는 이전 레이어에서 출력되는 특징맵을 인가받아 특징을 추출하여 각 레이어별 특징맵(A_k^C)을 획득한다. 신경망(120)은 다수의 특징맵(A_1^C, A_2^C, …, A_n^C)을 획득할 수 있다.

[0077] 각 레이어별 특징맵(A_k^C)이 획득되면, 클래스 분류부(123)가 미리 학습된 방식에 따라 최종 특징맵(A_n^C)이 기지정된 다수의 클래스 각각에 대응하는 수준을 계산하여 다수의 개별 클래스 분류 결과(y^c)로 구성된 클래스 분류 결과(y)를 획득한다(S40).

[0078] 이후 다수의 개별 클래스 분류 결과(y^c) 중 선택된 클래스에 대응하는 개별 클래스 분류 결과(y^c)를 초기 관련성(R^c)으로 설정한다(S50). 그리고 설정된 초기 관련성(R^c)을 신경망(120)의 최종 레이어로부터 역순으로 순차적으로 역전파하여, 레이어별 관련성 맵(R_k^c)을 획득한다(S60).

[0079] 관련성이 타겟 레이어(t)까지 전파되었는지 판별한다(S70). 만일 관련성이 타겟 레이어(t)까지 전파된 것으로

판별되면, 각 레이어에 대한 관련성 맵(R_k^c) 각각을 전역 평균 풀링하여, 각 레이어에 대한 관련성 가중치(α_k^c)를 획득한다(S80).

[0080] 각 레이어에 대한 관련성 가중치(α_k^c)가 획득되면, 레이어별 관련성 가중치(α_k^c)를 대응하는 레이어에서 출력된 특징맵(A_k^c)과 원소곱하여 가중하고, 특징맵(A_k^c)의 채널 방향으로 전역 평균 풀링하여 2차원으로 변환하여 각 레이어에 대한 관련성 가중 특징맵($\alpha_k^c A_k^c$))을 획득하고, 각 레이어에 대한 관련성 가중 특징맵($\alpha_k^c A_k^c$))을 합하여, 타겟 클래스(t) 및 타겟 레이어(k)에 대한 관련성 가중 클래스 활성맵($L_{Relevance-CAM}$)을 획득할 수 있다. 여기서 특징맵(A_k^c)의 채널 방향으로 전역 평균 풀링하는 것은 획득된 영상과 동일하게 2차원의 관련성 가중 클래스 활성맵($L_{Relevance-CAM}$)을 획득하기 위해서이며, 각 레이어에 대한 관련성 가중 특징맵($\alpha_k^c A_k^c$))는 동일한 2차원 크기로 변형될 수 있다.

[0081] 추가적으로 획득된 관련성 가중 클래스 활성맵($L_{Relevance-CAM}$)의 크기가 획득된 영상과 상이한 경우, 관련성 가중 클래스 활성맵($L_{Relevance-CAM}$)의 크기가 영상과 동일해지도록 업스케일링할 수 있다.

[0082] 본 발명에 따른 방법은 컴퓨터에서 실행시키기 위한 매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다. 여기서 컴퓨터 판독가능 매체는 컴퓨터에 의해 액세스 될 수 있는 임의의 가용 매체일 수 있고, 또한 컴퓨터 저장 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하며, ROM(판독 전용 메모리), RAM(랜덤 액세스 메모리), CD(컴팩트 디스크)-ROM, DVD(디지털 비디오 디스크)-ROM, 자기 테이프, 플로피 디스크, 광데이터 저장장치 등을 포함할 수 있다.

[0083] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다.

[0084] 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0085]

110: 영상 획득부 120: 신경망

121: 특징 추출부 123: 클래스 분류부

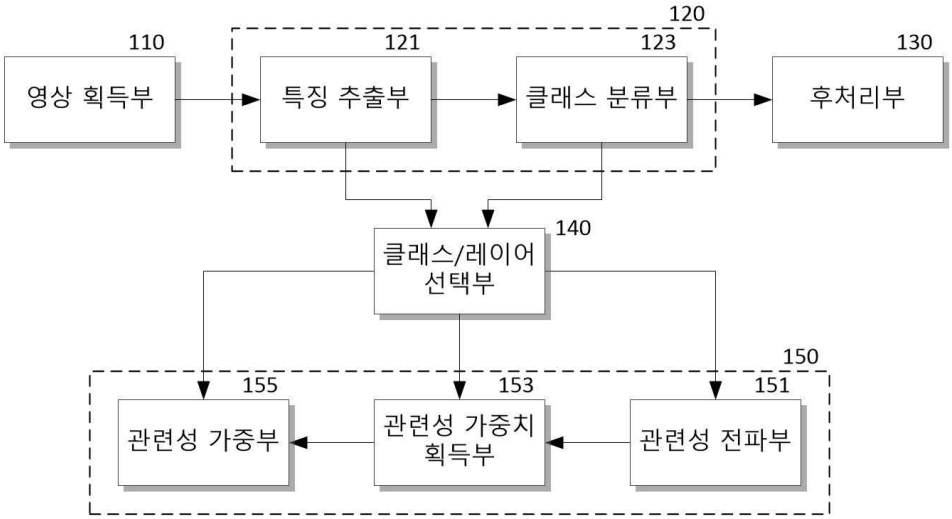
130: 후처리부 140: 클래스/레이어 선택부

150: 관련성 가중 CAM 획득부 151: 관련성 전파부

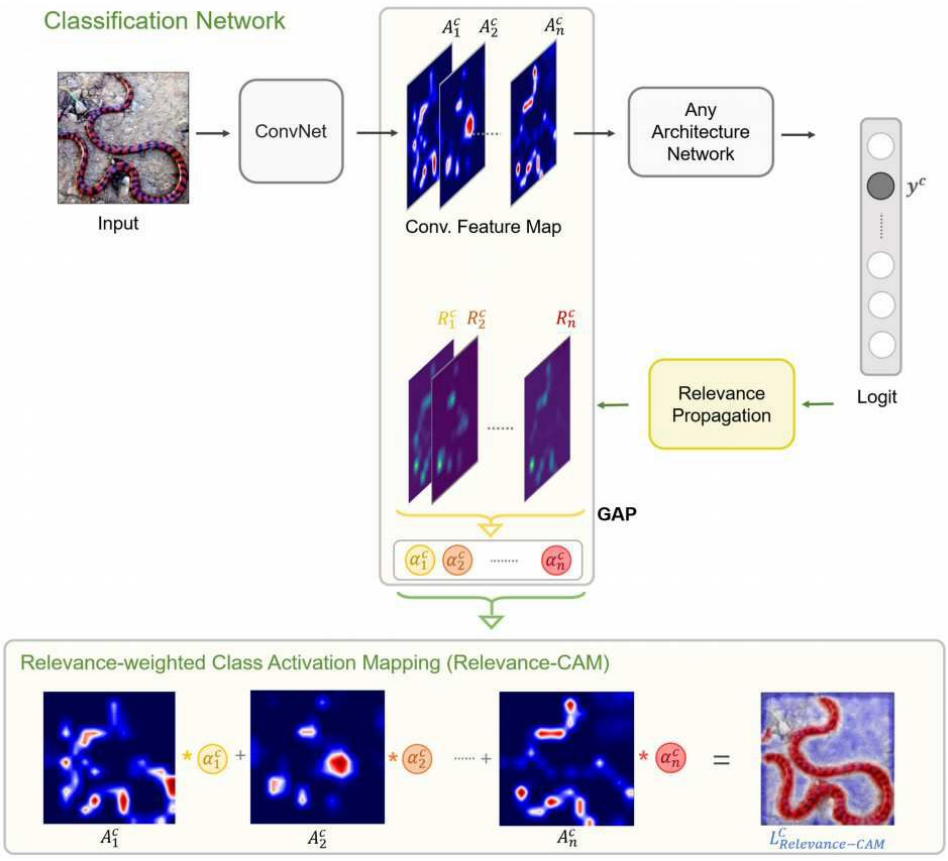
153: 관련성 가중치 획득부 155: 관련성 가중부

도면

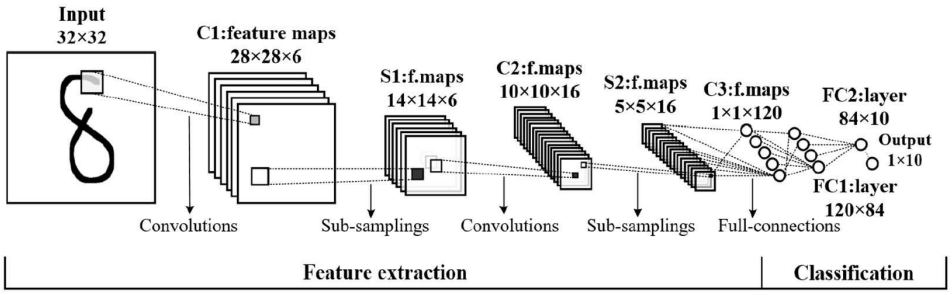
도면1


110
영상 획득부
120
121
특징 추출부
123
클래스 분류부
130
후처리부
140
클래스/레이어 선택부
150
155
관련성 가중부
153
관련성 가중치 획득부
151
관련성 전파부

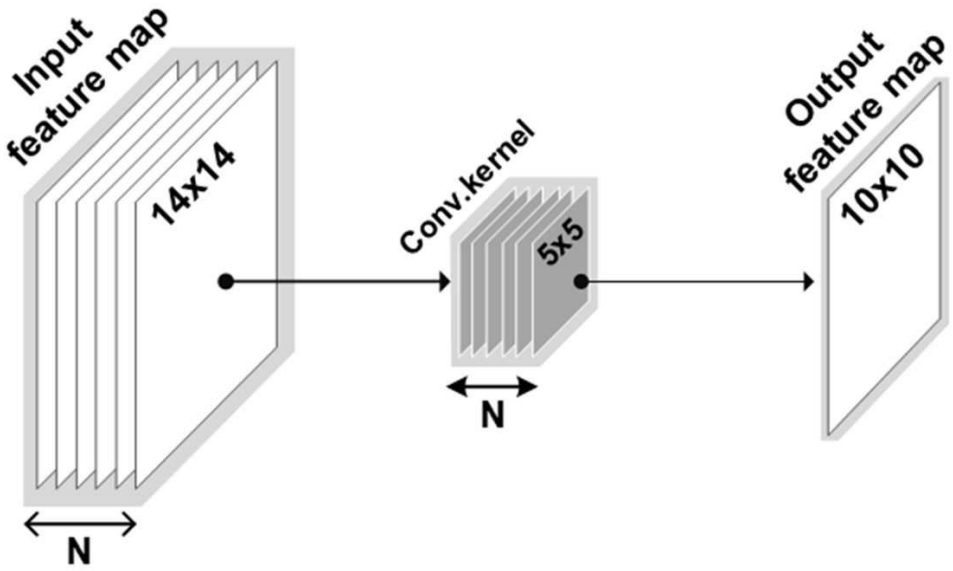
도면2


Classification Network
Input
ConvNet
Conv. Feature Map
Any Architecture Network
Logit
Relevance Propagation
GAP
Relevance-weighted Class Activation Mapping (Relevance-CAM)

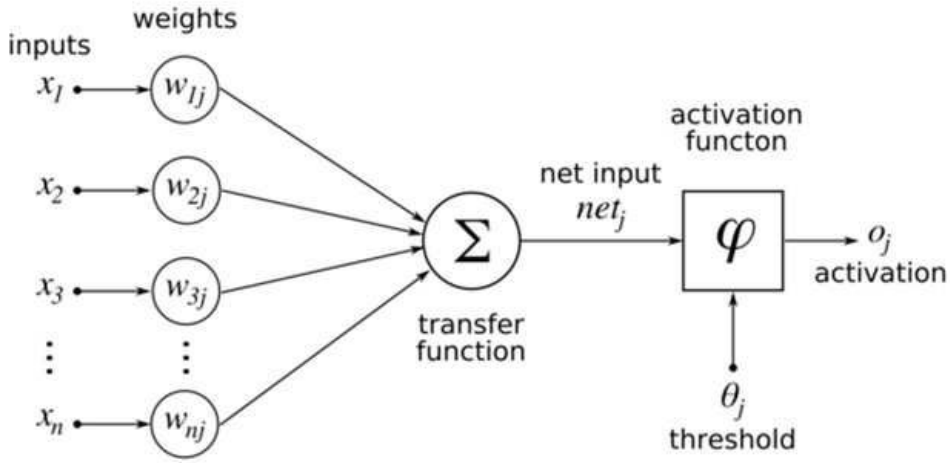
도면3


Input
32×32
C1:feature maps
28×28×6
S1:f.maps
14×14×6
C2:f.maps
10×10×16
S2:f.maps
5×5×16
C3:f.maps
1×1×120
FC2:layer
84×10
Output
1×10
FC1:layer
120×84
Convolutions
Sub-samplings
Convolutions
Sub-samplings
Full-connections
Feature extraction
Classification

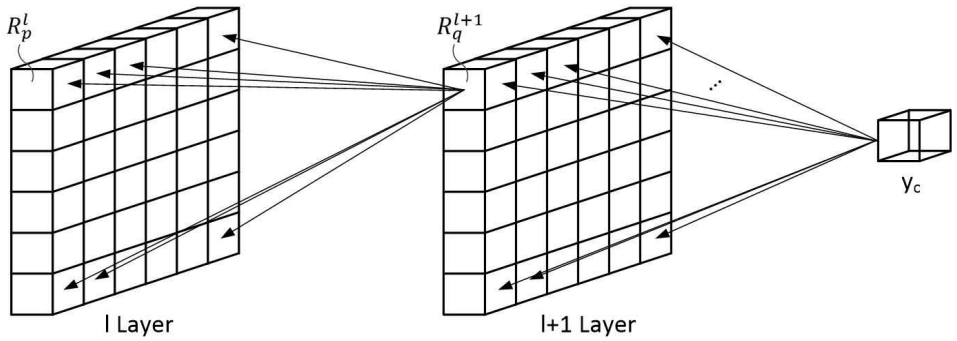
도면4


Input
feature map
14x14
N
Conv.kernel
5x5
N
Output
feature map
10x10

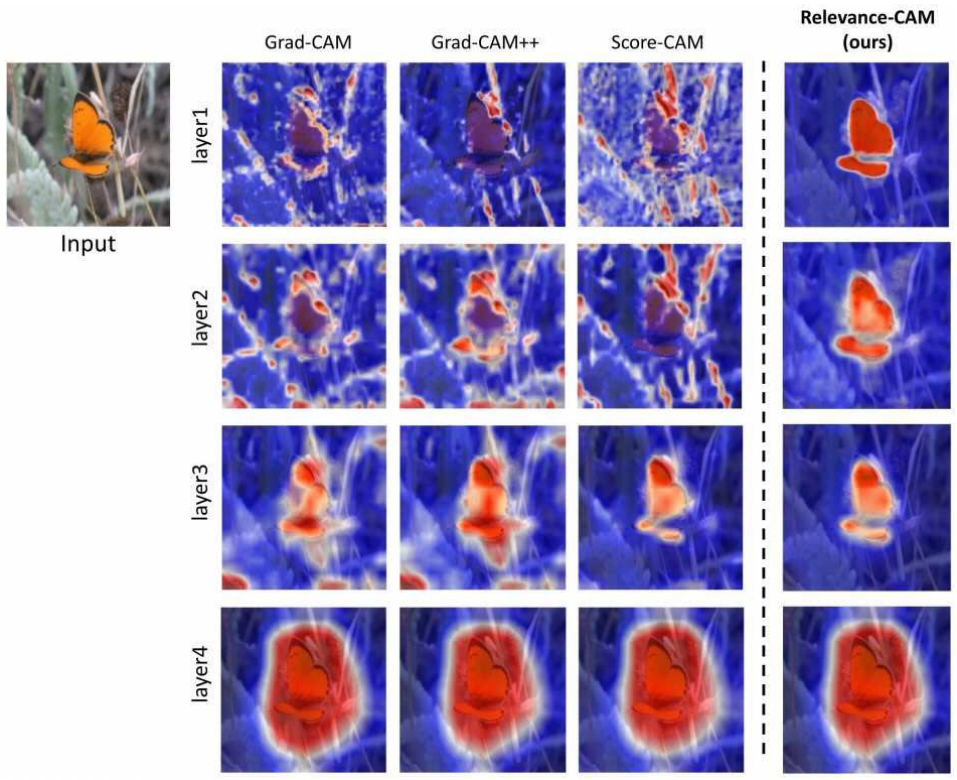
도면5


inputs
weights
x_1
x_2
x_3
x_n
w_{1j}
w_{2j}
w_{3j}
w_{nj}
Σ
transfer
function
net input
net_j
activation
functon
φ
o_j
activation
θ_j
threshold

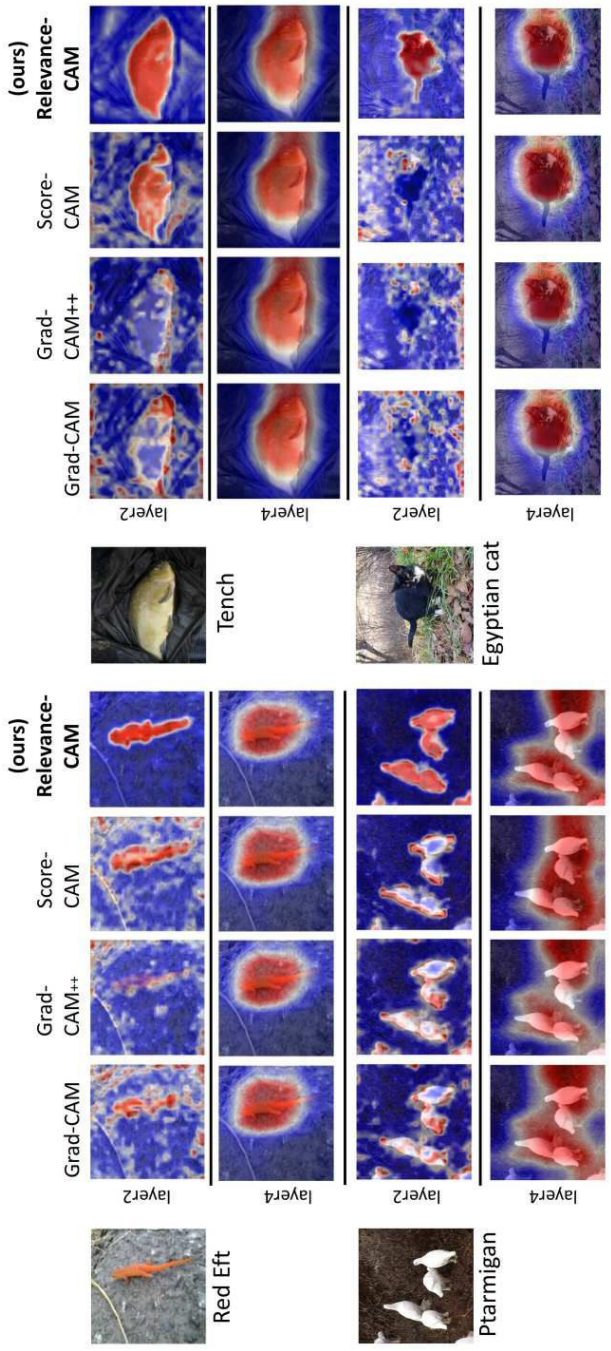
도면6


R_p^l
R_q^{l+1}
y_c
l Layer
l+1 Layer

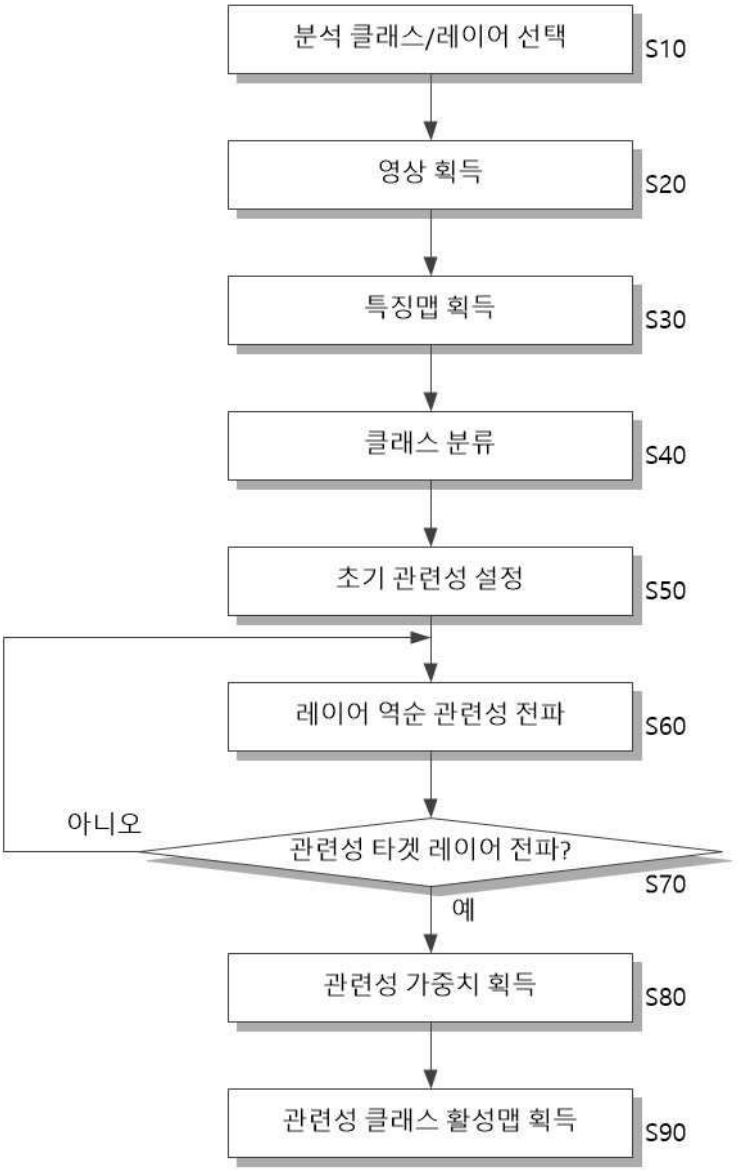
도면7


Grad-CAM
Grad-CAM++
Score-CAM
Relevance-CAM (ours)
Input
layer1
layer2
layer3
layer4

도면8


(ours) Relevance-CAM
Score-CAM
Grad-CAM++
Grad-CAM
layer2
layer4
layer2
layer4
Tench
Egyptian cat
(ours) Relevance-CAM
Score-CAM
Grad-CAM++
Grad-CAM
layer2
layer4
layer2
layer4
Red Eft
Ptarmigan

도면9


분석 클래스/레이어 선택
S10
영상 획득
S20
특징맵 획득
S30
클래스 분류
S40
초기 관련성 설정
S50
레이어 역순 관련성 전파
S60
아니오
관련성 타겟 레이어 전파?
S70
예
관련성 가중치 획득
S80
관련성 클래스 활성맵 획득
S90