

SISE 의 계산량 감소를 위한 Attribution mask Compress-SISE 기법 제안

김용호, 김지하, 박현희*
명지대학교

yhkim98@mju.ac.kr, yaki5896@mju.ac.kr, *hhpark@mju.ac.kr

Attribution mask Compress-SISE method for reducing the calculation volume of SISE

Yongho Kim, Jiha Kim, Hyunhee Park*
Myoungji Univ.

요 약

본 논문은 SISE 기법에서 유사한 속성 마스크들을 압축하여 설명맵 생성시 연산량을 줄이는 AC-SISE 기법을 제안한다. AC-SISE의 성능을 측정하기 위해 VGG16 모델로 실험한 결과 EBPG, mIoU에서 SISE 보다 각 1.40%, 0.07% 낮은 성능을 보이지만 속성 마스크는 17.47% 감소했다. ResNet50 모델은 EBPG에서 SISE 보다 1.90% 낮은 성능을 mIoU에서는 0.89% 높은 성능을 보이고 속성 마스크는 34.69% 더 감소한 성능을 보였다.

I. 서 론

최근 영상 처리 분야에서 딥러닝 모델의 결정을 설명하기 위해 설명맵(explanation map)을 생성하는 다양한 기법들이 연구되고 있다. 대표적인 기법으로 Randomized Input Sample for Explanation(RISE)[1]와 Semantic Input Sampling for Explanation(SISE)[2]이 있다. RISE는 원본 입력 이미지에 랜덤하게 생성한 마스크를 씌워 설명이 필요한 모델의 입력으로 넣은 후 모델의 출력 결과를 저장한다. 그 후 입력에 따라 저장된 출력 결과들을 비교하여 모델의 출력을 결정하는데 중요한 부분을 식별한다. 하지만 RISE는 이러한 과정을 약 8천번 진행해야 하기 때문에 연산량이 크고 랜덤 마스크를 사용하기 때문에 같은 입력 이미지에 대해서 설명맵을 생성할 때마다 결과가 달라진다는 단점이 있다. 이를 개선한 기법이 SISE다. SISE는 랜덤 마스크를 사용하는 대신 CNN 모델이 입력 이미지로부터 특징을 추출한 특징맵(feature map)을 마스크로 사용한다. 특징맵을 마스크로 만든 것을 속성 마스크(attribution mask)라고 한다. 이로 인한 장점은 RISE 보다 객체를 더 정확히 찾고 설명맵의 결과가 생성할 때마다 달라지지 않는다. 또한,

기존 랜덤 마스크의 수보다 CNN 모델에서 사용하는 특징맵의 수가 더 적기 때문에 연산량이 줄어든다.

본 논문은 SISE 기법에서 사용하는 속성 마스크를 압축하는 기법을 추가하여 설명맵 생성 시 필요한 연산량을 더 감소시키는 연구를 진행한다.

II. 본론

그림 1은 제안하는 기법인 Attribution mask Compress-SISE(AC-SISE)의 layer visualization map을 생성하는 과정을 도식화한 것이다. Layer visualization map은 개별의 convolution block의 pooling layer 이전 layer에서 다음과 같은 과정을 통해 생성한다.

우선 CNN 모델에 이미지를 입력하면 특징맵 추출 단계에서 각 layer 별로 특징맵을 추출한다. 추출한 특징맵들의 모든 픽셀에 대한 평균 gradient를 계산하는 단계가 있는데 이는 특징맵을 필터링 할 때 사용한다.

특징맵들을 필터링 하는 기준은 논문 [1]과 같이 특징맵 각 픽셀들의 gradient를 평균한 값이 0 이상인 특징맵들을 필터링한다. 본 연구에서는 추가적으로 특징맵 픽셀의 최대값이 0인 특징맵들 또한 필터링했다. 필터링 한 특징맵들은 post-processing 과정을 거친다.

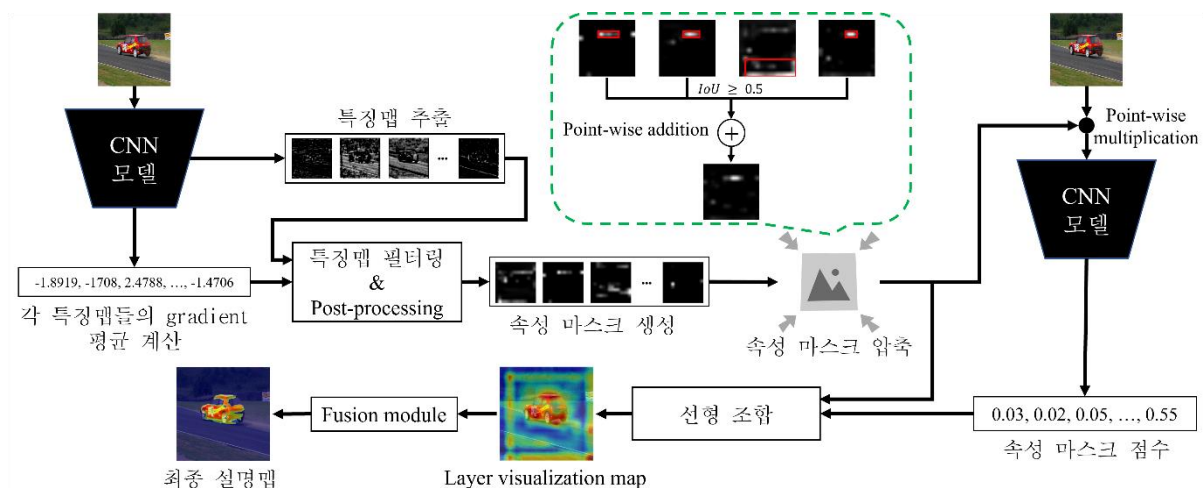


그림 1. AC-SISE 기법으로 layer에서 visualization map을 생성하는 과정

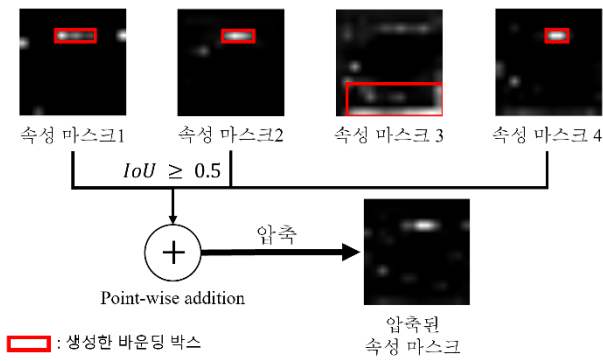


그림 2. 속성 마스크 압축 방법

Post-processing 과정에서는 속성 마스크의 크기를 입력 이미지 크기에 맞게 bilinear interpolation 알고리즘으로 upsampling 을 한 후 정규화를 한다.

다음 과정은 속성 마스크 압축이다. 속성 마스크 압축은 SISE 의 연산량을 줄이기 위해 본 연구에서 추가한 기법이다. 그림 2 는 같은 layer 에서 생성한 서로 다른 속성 마스크들을 압축하는 방법을 표현한 그림이다. 그림 2 의 속성 마스크 1,2,4 를 보면 속성 마스크들 중에 비슷한 특징을 가지는 마스크들이 존재하는 것을 확인할 수 있다. 비슷한 특징을 가지는 속성 마스크를 압축하는 방법은 다음과 같다. 생성한 모든 속성 마스크에 otsu 이진화 알고리즘[3]을 적용한 후 이진화 된 마스크에 바운딩 박스를 생성한다. 생성된 바운딩 박스 중에서는 가장 큰 바운딩 박스만을 사용한다. 그 후 모든 속성 마스크 바운딩 박스의 IoU 를 비교하여 0.5 이상인 속성 마스크들을 point-wise addition 연산을 하여 하나의 속성 마스크로 압축한다. 속성 마스크가 압축됨에 따라 모델을 통과한 속성 마스크의 점수를 계산하기 위한 수가 줄어들고, 그에 따라 layer visualization map 을 생성하기 위한 연산량과 시간이 감소한다.

위 과정들을 통해 각 layer 에서 생성한 layer visualization map 은 SISE 에서 사용한 fusion module[2]로 결합하여 최종적으로 설명맵을 생성한다.

III. 실험

실험에는 20 개의 클래스와 바운딩 박스 레이블이 존재하는 PASCAL VOC 2007[4] 데이터 셋을 사용했다. 사용 모델은 VGG16[5]과 ResNet50[6] 모델을 사용했다. 각 모델은 테스트 데이터 세트에서 89.77%, 85.82%의 top-5 정확도를 가진다.

두 기법의 성능을 비교하기 위한 성능 지표로는 Energy-Based Pointing Game(EBPG)[7], Mean Intersection over Union(mIoU), 특징맵 감소율, 설명맵 생성 시간을 사용한다.

EBPG 는 타겟 클래스 바운딩 박스내에 설명맵의 픽셀이 얼마나 많이 포함되는지를 나타내는 성능 지표이다. 먼저 타겟 클래스 바운딩 박스 안밖으로 입력 이미지를 이진화 하고 바운딩 박스 내부 영역은 1, 외부 영역은 0 으로 할당한다. 그 후, 생성한 설명맵과 point-wise multiplication 연산한 다음 타겟 클래스의 바운딩 박스 내에 존재하는 설명맵의 픽셀 수를 설명맵의 전체 픽셀수로 나누어 계산한다.

mIoU 는 두 바운딩 박스가 얼마나 일치하는지를 계산하는 지표인 IoU 값의 평균을 계산한 지표이다. 본 논문의 실험에서는 전체 테스트 데이터 세트에 대해서 정답 바운딩 박스와 설명맵으로 생성한 바운딩 박스의 IoU 를 계산한 후 평균을 계산했다. 설명맵의 바운딩 박스는 설명맵 최댓값의 20%를 넘는 픽셀들을 선택 후,

선택된 픽셀들이 가장 많이 포함되는 바운딩 박스를 선택해 생성한다.

실험 결과 VGG16 모델에서는 SISE 와 AC-SISE 의 EBPG 가 각각 46.95%, 45.55%로 AC-SISE 가 1.40% 낮고 mIoU 또한 각각 13.50%, 13.42%로 AC-SISE 가 0.07% 더 낮다. 하지만 AC-SISE 의 특징맵 감소율이 -67.34%로 SISE 보다 17.47% 더 감소했다.

ResNet50 모델에서는 SISE 와 AC-SISE 의 EBPG 가 각각 50.24%, 48.34%로 AC-SISE 가 1.90% 낮고 mIoU 는 각각 10.20%, 11.09%로 AC-SISE 가 0.89% 더 높은 성능을 보였다. 특징맵 감소율은 -83.54%로 SISE 보다 34.69% 더 감소했다.

표 1 실험 결과

모델	성능 지표	SISE	AC-SISE
VGG16	EBPG (%)	46.95	45.55
	mIoU (%)	13.50	13.42
	특징맵 감소율 (%)	-49.87	-67.34
	설명맵 생성 시간 (초)	2.3210	1.7451
ResNet-50	EBPG (%)	50.24	48.34
	mIoU (%)	10.20	11.09
	특징맵 감소율 (%)	-48.85	-83.54
	설명맵 생성 시간 (초)	5.6108	2.0209

IV. 결론

본 논문에서는 SISE 의 속성 마스크를 압축하는 기법인 AC-SISE 를 제안했다. 실험 결과 SISE 의 속성 마스크를 압축했을 때 모델 설명 성능이 크게 낮아지지 않으면서 연산량은 크게 감소된 것을 확인할 수 있다.

향후 연구에서는 속성 마스크 압축 기법을 다양한 방식으로 연구하여 성능을 개선할 계획이다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00990, 설명가능한 인공지능 기반 무선랜 네트워크 시스템 고도화 핵심 기술 연구, No. 2021-0-00368, 6G 서비스를 위한 인공지능/머신러닝 기반 자율형 MAC 개발)

참 고 문 헌

- [1] Petsiuk, Vitali, Abir Das, and Kate Saenko, "Rise: Randomized input sampling for explanation of black-box models," arXiv preprint arXiv:1806.07421, 2018.
- [2] Sattarzadeh, Sam, et al, "Explaining convolutional neural networks through attribution-based input sampling and block-wise feature aggregation," Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, No. 13, 2021.
- [3] Otsu, Nobuyuki, "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics, Vol. 9, No. 1, p. 62-66, 1979.
- [4] Everingham, Mark, et al, "The pascal visual object classes (voc) challenge," International journal of computer vision, Vol. 88, No. 2, p. 303-338, 2010.
- [5] Simonyan, Karen, and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [6] He, Kaiming, et al, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, p. 770-778, 2016.
- [7] Wang, Haofan, et al, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, p. 24-25, 2020.