설명 가능한 이미지 인식을 위한 채널 주의 기반 딥러닝 방법

백나 1, 조인휘 2
1 한양대학교 컴퓨터소프트웨어학과 석사과정
2 한양대학교 컴퓨터소프트웨어학과 교수

bnn0118@hanyang.ac.kr, iwjoe@hanyang.ac.kr

Deep Learning Methods for Explainable Image Recognition

BaiNa[1], Inwhee Joe[2]
1 Department of Computer Software, Master's Course, Hanyang University
2Professor, Department of Computer Software, Hanyang University

요 약

본 실험 연구에서는 주의 메커니즘과 컨볼루션 신경망을 결합하여 모델을 개선하는 방법을 탐색하는 딥 러닝 기술을 소개한다. 이 기술은 지도 학습 방식을 위해 공개 데이터 세트의 쓰레기 분류 데이터를 사용하고, Grad-CAM 기술과 채널 주의 메커니즘 SE 를 적용하여 모델의 분류 의사 결정 과정을 더 잘 이해하기 위해 히트 맵을 생성한다. Grad-CAM 기술을 사용하여 히트 맵을 생성하면 분류 중에 모델이 집중하는 영역을 시각화할 수 있다. 이는 모델의 분류 결정을 설명하는 방법을 제공하여 다양한 이미지 카테고리에 대한 모델 결정의 기초를 더 잘 이해할 수 있다. 실험 결과는 전통적인 합성곱 신경망과 비교하여 제안한 방법이 쓰레기 분류 작업에서 더나은 성능을 달성한다는 것을 보여준다. 주의 메커니즘과 히트맵 해석을 결합함으로써 우리 모델은분류 정확도를 향상시킬 수 있다. 이는 실제 응용 분야의 이미지 분류 작업에 큰 의미가 있으며 해석 가능성에 대한 딥 러닝 연구 진행을 촉진하는 데 도움이 된다.

1. 서론

이미지 분류는 컴퓨터 비전 분야의 핵심 작업으로, 많은 실제 응용 분야에서 폭넓게 적용된다. 그러나 딥 러닝 모델이 개발되고 널리 적용되면서 몇 가지 중요한 문제도 표면화되었다. 그 중 두 가지 주요 문제는 모델 해석 불가능성과 분류 오류이다.

딥 러닝 모델은 대량의 데이터에서 기능을 자동으로 학습하고 많은 작업에서 뛰어난 성능을 달성할 수 있다는 점에서 강력합니다. 그러나 이러한 "블랙박스" 특성은 모델의 해석 불가능성이라는 중요한 문제도 야기합니다. 기존의 기계 학습 알고리즘은 해석 가능한 모델을 제공할 수 있는 경우가 많지만 딥 러닝 모델의 복잡성으로 인해 이해하기가 어렵다 [1]. 딥러닝 모델을 더 잘 적용하려면 해석 가능성을 개선하여 의사 결정 과정을 투명하게 만드는 것이 시급한다.

이러한 과제를 해결하기 위해 이 프로젝트에서는 딥 러닝이미지 분류의 해석 불가능성 및 잘못된 분류 문제를 해결하는 것을 목표로 주의 메커니즘을 도입한다. 어텐션 메커니즘 SE 를 도입함으로써 모델의 성능을 향상시키고, 해석하기 쉽게 만들고, 컴퓨터비전 분야에 더 많은 응용 가능성을 제공할 것으로 기대한다.

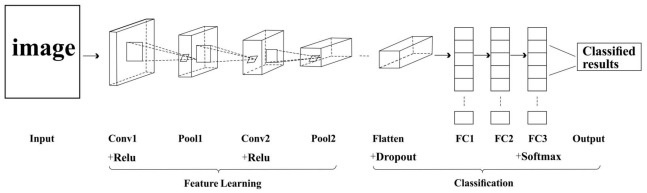
2. 방법

2.1 컨볼루션 신경망

본 연구에서는 ResNet-50 네트워크에서 개선된 컨볼루션 신경망 구조를 사용하였다 [2]. 연구의 초기단계에서 우리는 다양한 구조의 컨볼루션 신경망 모델의 성능을 비교하기 위해 작은 샘플 데이터를 사용했습니다.기본적으로 모델의 정확도를 손상시키지 않는다는 전제 하에 비교적 간단한 컨볼루션 신경망 구조를 선택했다.

그림 1 에서 보는 바와 같이 본 연구에 사용된 컨볼루션 신경망 구조는 2 개의 컨볼루션 층, 2 개의 풀링 층 및 3 개의 완전 연결 층을 포함한다.훈련 과정

에서 손실 함수의 변화에 따라 반복 횟수를 20 회, 배치 크기를 64 로 설정하고 최적의 학습률을 0.001 로 결정했다. 동시에 무작위 기울기 하강법을 사용하여 모델의 하이퍼 파라미터를 최적화하여 최적의 모델 구성을 결정한다. 모델의 전반적인 성능을평가하기 위해 테스트 세트를 사용했다. 테스트 세트의 각이미지에 대해 해당 예측 점수를 얻을 수 있으며, 이점수는 모델의 쓰레기 유형 인식 정확도를 나타나다 [3].

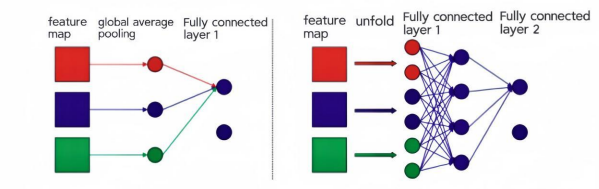


(그림 1) 컨볼루션 신경망 구조도.

2.2 경사 가중 클래스 활성화 히트맵

딥러닝 모델을 구축하는 과정에서 모델이 판단을 내리기위해 주로 어떤 정보나 특징에 의존하는지 이해할 수 없기때문에 딥러닝 모델을 설명할 수 없고 '블랙박스' 현상이 발생한다. 딥러닝 모델의 연산 과정에 대한 "블랙박스" 문제에 대응하기 위해, 연구자들은 혁신적인 해석 방법을 많이 제안해왔다. 그중에서도 Gradient class activation map (Grad-CAM)은 대표적인 딥러닝 결과 시각화 및 해석 기술입니다 [4][9] .Grad-CAM 은 CAM(Class Activation Map)을 기반으로 개발되었다.

그림 2 에서 볼 수 있듯이 CAM 의 기본 원리는 CNN 의 마지막 컨볼루션에 의해 출력된 특징 맵이 채널의 가중치 중첩 후 활성화 값이 있는 영역이 이미지의 객체가 있는 영역이 되는 것이다. 그리고 이것이 중첩된 단일 채널 특징 맵이 입력 이미지에 중첩되면 이미지에서 객체가 위치한 영역이 강조 표시될 수 있으며, 모델이 어떤 영역을 사용하는지 관찰하기위해 모델 특징 맵을 시각화하는 데 사용할 수 있다. 이미지 카테고리를 예측한다. 그러나 CAM 의 적용에는 몇가지 제한 사항이 있다. 이는 전역 평균 풀링 계층과 단 하나의 완전 연결 계층이 있는 신경망 모델에만 적용 가능하며 여러 완전 연결 계층이 있는신경망에는 적용할 수 없다. 따라서 우리는 모델을 처리하기 위해 Grad-CAM 알고리즘을 선택했다.

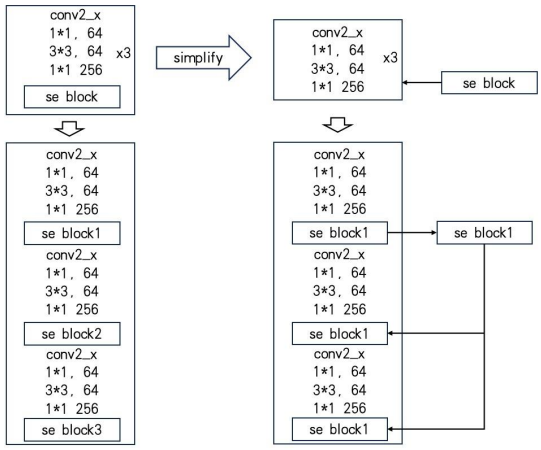


(그림 2) CAM 과 Grad-CAM 의 핵심 원리 비교 설명도.

3. 주의 메커니즘 기반의 ResNet50

3.1 주의 메커니즘 개선

개선된 주의 메커니즘의 실현은 네트워크에 여러 개의 글로벌 텐서를 설정하는 것이며 텐서의 수는 네트워크 모델의 모듈 수와 동일하며 그 역할은 해당 모듈의 주의력을저장하고 모듈의 첫 번째 주기에서 주의 메커니즘을 학습하고 해당 텐서에 학습 결과를 저장하는 것이다. 네트워크 모듈 뒤의 주기에서 새로운 주의 메커니즘을 학습하지 않고 해당 모듈의 텐서를 호출하여 개선 목적을 달성한다 [5]. 구조는 그림 3 과 같다.



(그림 3) 주의 메커니즘 구조 다이어그램.

3.2 개선된 주의 메커니즘 기반의 ResNet50 모듈

주의 메커니즘의 사용은 네트워크에 플러그인으로 가입하고 훈련하는 것이다. 현재 비교적 성숙한 ResNet50 네트워크를 백본 기능 추출 네트워크로 선택하고 채널 주의 기반 ResNet50 네트워크(약칭 se_ResNet50)를 추가하여 ResNet50 네트워크를 추가하는 주의 메커니즘을 선택한다.

개선된 주의 메커니즘을 기반으로 개선된 채널 주의력(약칭 s_se_ResNet50)을 제안하며, 특징 추출 백본 네트워크는 ResNet50 이며, 이 네트워크는 5 개의 스테이지로 구성되며, 이 중 후자의 4 개의 스테이지 내부 모듈 구조는 유사하며 출력된 특징 맵의 채널 수와 스테이지 내부 모듈 사이클 횟수만 다르다.주의 메커니즘을 추가한 후 주의 모듈 개수는 스테이지내부의 모듈 사이클 횟수에 따라 증가한다. 그리고 본 논문에서는 각 모듈이 동일한 구조를 가지며 요구되는 주의력이 일치하는 주기 때문에 각 모듈에 주의력을 추가할 필요가 없으며, 단지 각 스테이지에 주의력을 추가하고 스테이지 내의 모든 모듈이 이 주의력을 공유하면 된다고 주장한다. 이렇게 하면 네트워크 학습의 부담을줄이고 네트워크 학습의 속도를 높일 수 있으며 과도한 주의로 인한 판단 간섭을 줄여 네

트워크의 정확도를 높일 수있다.

4. 구체적인 실험 절차

4.1 데이터 세트

(1) 데이터셋의 수집

데이터 세트의 획득은 훈련 모델이 충분한 다양성과 대표성을 갖도록 하는 핵심 단계이다.학습 작업을 감독하기위해 다양한 스팸 이미지를 포함하는 쓰레기 분류 데이터 Garbage Classification Data 를 선택했다.

(2) 이미지 전처리

본 실험에 사용된 데이터 전처리 방법에는 이미지 크기 정규화 및 이미지 향상이 포함된다. 이미지 크기 정규화는 데이터 세트에 포함된 이미지의 크기를 통일하는 것으로, 본 실험의 데이터 세트에 포함된 이미지가 다양하므로 이미지의 크기도 서로 다르다. 크기를 정규화해야 한다 [6]. 본 실험에서는 데이터 향상을 위해 이미지 뒤집기 방법을 사용한다.

4.2 이미지의 특징 벡터를 추출한다

첫째, 사전 훈련된 잔차 네트워크 ResNet50 을 사용하여 이미지의 심층 특징을 추출한다.둘째, DCT 변환을 사용하여 특성 정보를 추가로 압축하여 DCT 계수 행렬을 얻다.마지막으로 DCT 계수 행렬의 저주파 정보 영역에서 64 비트 계수를 캡처하고 감지 해시 알고리즘과 결합하여 고유 벡터 F 를 생성한다 [7].

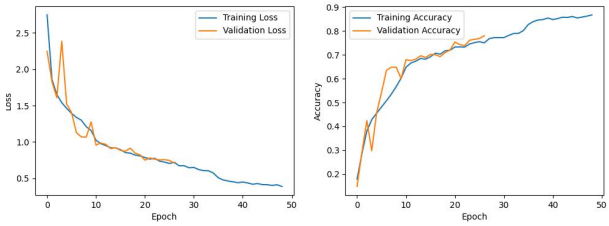
4.3 훈련 과정

본 논문에서는 제로부터 훈련하는 방법을 사용하였으며, 동일한 데이터 세트를 사용하여 다중 분류 문제를 다루었다. 훈련 시 배치 크기는 128 이었고, 네트워크의 기본 입력 이미지 크기는 128x128 이었다. 각 네트워크는 100 개의 에포크 동안 훈련되었으며, 초기 학습률은 0.001 이었다. Adam 최적화 알고리즘을 사용하였으며, 손실 함수로는 교차 엔트로피 손실 함수를 사용했다. Adam 최적화 알고리즘은 GradCam 과 주의 메커니즘의 장점을 결합하여그레이디언트의 일차 모멘트와 이차 모멘트를 종합적으로 고려하여 업데이트 단계의 크기를 계산한다 [8].

4.4 실험 결과

이 섹션에서는 모델의 훈련 프로세스와 성능을 평가하기위해 개선 전후의 ResNet50 모델의 훈련 손실 및 검증 손실, 훈련 정확도 및 검증 정확도 곡선에 대한 포괄적인 분석을 수행한다.

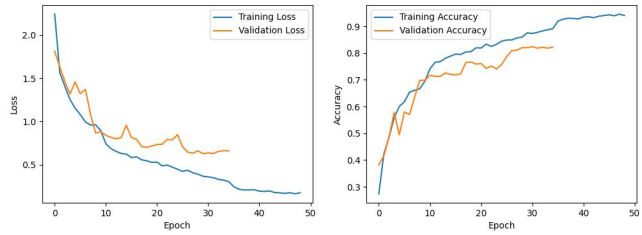
그림 4 는 각각 ResNet50 모델의 SE 주의 메커니즘이 추가되지 않은 훈련 및 검증 손실 그래프와 훈련 및 검증 정확도 그래프이다. 그림 4 의 왼쪽은 50 개의 epoch 에서 '훈련 손실'과 '검증 손실'을 나타내고, 그림 4 의 오른쪽은 50 개의 epoch 에서 '훈련 정확도'와 '검증 정확도'를 나타나다. 그림에서 알 수 있듯이 epoch 가 증가할수록 훈련 손실과 검증 손실은 감소하고 훈련 정확도와 검증 정확도는 모두 증가함을 알 수 있다. 이는 모델이 예측을 학습하고 개선한다는 것을 의미한다.



(그림 4) ResNet50 모델은 SE 모듈의 훈련 및 검증 손실 그래프, 훈련 및 검증 정확도 그래프를 추가하지 않다.

결과는 모델이 50 개의 epochs 훈련 후 테스트 세트에서 약 77.59%의 정확도에 도달했음을 보여준다. 또한 모델의훈련 정확도와 검증 정확도 곡선을 관찰하고 둘 사이의 차이가 작다는 점에 주목한다. 이는모델이 훈련 데이터와 검증 데이터의 성능이 비교적 일관되고 일반화 능력이 우수함을 나타나다. 훈련 과정에서 훈련 정확도 곡선이 꾸준히 향상되었으며 검증 정확도 곡선도 유사한 경향을 보였다. 이것은 모델이 훈련 중에 보이지 않는 검증 데이터로효과적으로 학습하고 일반화할 수 있음을 시사한다. 훈련과 검증 정확도 간의 격차를 줄임으로써 ResNet50 모델은 입력 데이터에 대한 우수한 이해와 일반화 능력을보여 이미지 분류 작업에서 우수한 성능을 추가로 검증한다.

그림 5 는 SE 주의 메커니즘이 추가된 ResNet50 모델의 훈련 및 검증 손실 곡선과 훈련 및 검증 정확도 곡선을보여준다. 그림에서 볼 수 있듯이 epoch 가 증가할수록 훈련 손실과 검증 손실은 모두 감소하는 반면, 훈련 정확도와검증 정확도는 모두 증가한다.



(그림 5) SE 모듈이 추가된 ResNet50 모델의 훈련 및 검증 손실 곡선, 훈련 및 검증 정확도 곡선.

정확도 곡선을 관찰하면 약 30 epoch 이후 훈련 정확도와검증 정확도 곡선이 점차 평준화된다. 이는 모델이 특정 수준의 예측 정확도에 도달했으며 추가 교육을 통해 모델 성능이 크게 향상되지 않을 수 있음을 의미한다. 결과는 50 번의 훈련 후에 우리 모델이 테스트 세트에서 약 83.33%의 정확도를 달성했다는

것을 보여준다. 훈련 세트와 검증 세트의 정확도에는 약간의 차이가 있지만 그 차이는 크지 않다. 이는 모델이 심각한 과적합 없이 훈련 세트와검증 세트 모두에서 유사한 예측 정확도를 달성했음을 보여준다.

SE 어텐션 메커니즘이 있는 경우와 없는 경우의 두 훈련프로세스의 손실 및 정확도 곡선을 비교함으로써 의미 있는 실험 결과를 관찰했다. SE 주의 메커니즘이 없는 모델은 77.59%의 정확도를 달성했으며, SE 주의 메커니즘을 추가하여 모델 검증 정확도가 효과적으로 향상되어 최종적으로 83.33%의 정확도에 도달했다. 우리의 결과는 SE 주의 메커니즘을 추가하는 것이 모델 성능을 향상시키는 데 중요하다는 것을 보여준다.

최종적으로 우리의 실험 결과는 아래 그림과 같으며, 그림 6 은 SE 모듈을 추가하고 그림 7 는 SE 모듈을 추가하지않은 것이다.

(그림 6) ResNet50 모델에 SE 모듈을 추가하는 히트맵.

(그림 7) ResNet50 모델에 SE 모듈을 추가하지 않은 히트맵.

주의 메커니즘(SE 모듈)과 주의 메커니즘이 추가되지않은 열적 시도를 비교 분석하여 다음과 같은 결론을 도출하였다.

첫째, 주의 메커니즘(SE 모듈)의 도입은 쓰레기 분류 모델의 성능을 크게 향상시킨다. 열적 노력을 비교하면 SE 모듈을 추가한 후 모델이 쓰레기 분류 작업과 밀접하게 관련된 영역에 더 명확하게 초점을 맞추고 강조할 수 있음을관찰할 수 있다. 이것은 주의 메커니즘이 모델이 중요한특성 정보를 더 잘 포착하고 활용하는 데 도움이 되어 모델의 분류 정확도와견고성을 향상시킬 수 있음을 시사한다.

둘째, 주의 메커니즘을 추가하지 않은 히트 맵은 모델이 이미지의 다양한 영역에 명백히 차별적인 주의를 기울이지않음을 보여준다. 대조적으로, 주의 메커니즘을 추가한 후의 히트 맵은 쓰레기 분류 작업에 더 중요한 관심 영역을 더 분명하게 보여준다. 이는 주요 특징에 집중하고 추출하는 모델의 능력을 향상시키는 데 있어 어텐션 메커니즘의 효율성을 추가로 검증한다.

5. 결론

주의 메커니즘(SE 모듈)이 추가된 쓰레기 분류 모델은 열적 노력 비교 결과에서 명백한 이점을 보여준다. 주의 메커니즘을 도입함으로써 모델은 쓰레기 분류 작업과 관련된이미지 특성에 더 효과적으로 초점을 맞추고 활용할 수 있어 분류 정확도를 향상시킬수 있다. 이것은 쓰레기 분류 분야에서 주의 메커니즘의 추가 연구와 적용을 위한 중요한 지침과 시사점을 제공한다.

요약하면, 이 실험을 통해 GradCam 과 결합된 개선된 Resnet50 모델이 해석 가능한 이미지 분류에서 컨볼루션 신경망을 개선하는 데 명백한 이점이 있음을 명확히 하고 해석가능한 이미지 분류의 정확도를 크게 향상시키고 과적합이용이하여 우리 생활 환경 위생 부서에서 쓰레기 분류에 큰참고 역할을 하고 지능화를 촉진하는 데 도움이 된다. 앞으로 우리는 resnet50 의 해석 가능한 이미지 분류가 하루 빨리 정착될 날을 기대하기 위해 계속 노력할 것이다.

참고문헌

[1] Samek W, Wiegand T, M Uller K R. Explainable artificial intelligence:Understanding, visualizing and interpreting deep learning models[J].arXiv preprint arXiv:1708.08296, 2017.

[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.

[3] Glorot X, Bengio Y. Understanding the difficulty of training deep feed-forward neural networks[C]//Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2010: 249-256.

[4] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J].Advances in neural information processing systems, 2017, 30.

[6] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of big data, 2019, 6(1): 1-48.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recog- nition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[8] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[9] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps[J].arXiv preprint arXiv:1312.6034, 2013.