

이미지 이진 분류를 위한 CNN 과 바운딩 박스 회귀의 적용

조영채¹, 이현주², 이재호*
덕성여자대학교

¹20220691@duksung.ac.kr, ²20221004@duksung.ac.kr, *izeho@duksung.ac.kr

Application of CNN and Bounding Box Regression For Image Binary Classification

Youngchae Cho, Hyeonjoo Lee, Jaeho Lee
Duksung Women`s University

요 약

인공지능 기술 중 하나인 이미지나 영상 데이터 처리 분야는 빠르게 발전해왔다. 이미지 처리에는 고해상도의 이미지, 다양한 각도에서 본 이미지 등 여러 변수가 존재하기 때문에 이미지 인식 모델은 더 정교해야 할 필요가 있다. CNN 만으로 이미지의 이진 분류를 수행하기에는 여러 단점들이 존재하기에, 바운딩 박스 회귀 기법을 사용하여 보다 정교한 이미지 분류를 수행하였다. 이미지 전체가 아닌 바운딩 박스 처리한 부분을 중점적으로 학습한다는 점에서 더욱 효과적인 분류를 수행할 수 있게 된다.

I. 서 론

논문에서는 최근, 딥러닝 기술은 이미지 분석 등의 분야에서 상당한 발전을 이루어 냈다. 특히, 합성곱 신경망을 의미하는 CNN(Convolution Neural Network)은 이미지나 영상 데이터를 처리할 때 뛰어난 성능을 보여 널리 사용되고 있다. CNN 은 이미지의 특징을 효과적으로 추출할 수 있기에, 이미지의 분류에 많이 사용된다.

CNN 만을 사용하여 본인이 있는 사진과 없는 사진을 분류하는 모델도 보편적이다. 그러나 이에 한계가 존재하기 때문에, 바운딩 박스 회귀를 기반으로 한 이미지 분류 모델을 도입한다. 이는 배경이나 다른 객체에 영향 받지 않고 대상을 중점적으로 학습하기에 정확도를 높일 수 있다.

최종적으로, 본 논문에서는 바운딩 박스 예측 후 이미지를 이진 분류하는 모델을 제시한다. 이를 통해 개인의 얼굴을 정확하게 감지하고 분류함으로써 보다 정교한 이미지 분류를 실현할 수 있다.

II. 연구배경

2.1 CNN

CNN 은 합성곱 신경망을 뜻하는 말로, 이미지의 특징 추출을 통한 학습에 매우 효과적이다. 사람이 직접 설계한 특징에 의존하던 기존 알고리즘들과는 달리 학습에 의한 특징 추출이 더 효과적이라는 생각에서 착안하였다. CNN 은 합성곱 계층에서는 가중치를 포함한 필터를 통해 합성곱 연산을 하며 특징을 추출한 Feature

Map 을 얻는다. 그리고 Pooling 과, Padding 을 통해 사이즈를 조절하며 학습을 수행한다.^[1] 이렇게 이미지로부터 특징을 더욱 수월하게 추출할 수 있게 된다.

2.2 CNN 기반 이미지 분류 모델

먼저, 합성곱과 Pooling, Padding 을 반복하여 이미지의 이진 분류를 수행하였다. 그 결과, 찾으려는 대상이 있는 사진은 잘 판별하였다. 하지만 자신이 아닌 다른 사람의 얼굴도 나로 판별하는 점, 얼굴이 멀리 있거나 다른 각도로 있으면 잘 판별하지 못하는 점의 한계가 존재했다. 따라서 바운딩 박스 회귀 기법을 도입하여 얼굴을 더 중심으로 하여 모델을 개선하기로 하였다.

2.3 Bounding Box

바운딩 박스 (Bounding Box)는 이미지나 영상에서 특정 객체나 영역의 위치를 특정하기 위해 사용한다. 사용자가 지정한 레이블에 따라 바운딩 박스가 나타내는 바를 라벨링할 수 있다. 일반적으로 네 개의 값을 가지며, 좌표와 그의 크기로 표현된다.^[2]

이는 방대한 크기의 데이터에서 객체를 올바르게 탐지하기 위해 사용된다. 전체 이미지에서 특정 영역만 대상이 되기 때문에, 효율적으로 학습할 수 있다.

2.4 K-folds cross validation

교차검증 (K-folds Cross Validation)은 데이터를 K 개의 fold 로 만들어서 모델을 평가하는 기술이다. 각 fold 에서 한 fold 를 검증 데이터셋으로 선택하고 나머지 fold 들은 훈련 데이터셋으로 사용하여 모델을 훈련하고

평가한다.^[3] 이는 특정 데이터셋에 대한 과적합을 방지할 수 있다. 또한 더욱 일반화된 모델을 생성할 수 있다.

2.5 Bounding Box Regression

Bounding Box Regression (바운딩 박스 회귀)는 사용자가 직접 라벨링한 바운딩 박스의 좌표를 기반으로 그 바운딩 박스 좌표 내의 이미지를 학습한다. 그 후 주어진 사진에서 학습했던 바운딩 박스와 비슷한 부분을 예측하여 그 부분을 바운딩 박스로 예측한다.

III. 모델 설계 및 구성

3.1 Dataset

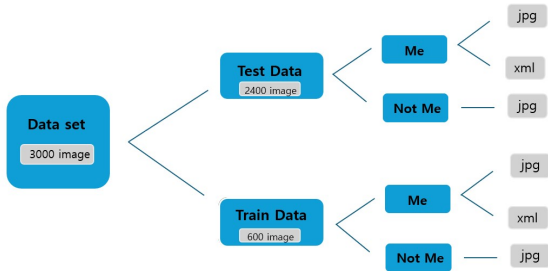


그림 1. 데이터 분류 및 전처리 형식

3.1.1 Data 형식

이 모델에서 가장 중심이 되는 이미지 데이터는 jpg 형식의 이미지 훈련 세트 2,400 장, 테스트 세트 600 장으로 구성된다. 이미지의 라벨링은 me / not me 라는 이름의 두 레이블로 나누어 절반씩 할당하였다. 따라서 그림 1 과 같이 훈련 세트와 테스트세트에는 각각 me 에 해당하는 사진과 not me 에 해당하는 사진들로 데이터가 구성된다.

3.1.2 Data 전처리

아이폰의 특성상 heic 형식의 파일로 저장되기 때문에 heic2png 라이브러리를 사용해 jpg 형식으로 변환했다. 이미지 데이터의 크기를 통일시키기 위해 opencv 라이브러리를 사용해 800 x 800 의 사이즈로 잘라주었다. 바운딩 박스의 라벨링은 얼굴을 바운딩 박스 처리하여 me 라는 레이블의 이름의 네 좌표가 저장되는 형식으로 진행했다. 모든 바운딩 박스 데이터는 xml 형식으로 저장되었다. 모든 데이터는 학습 수행시 Numpy 라이브러리를 사용해 numpy 배열로 변환 후 모델에 입력했다.

3.2 모델 알고리즘

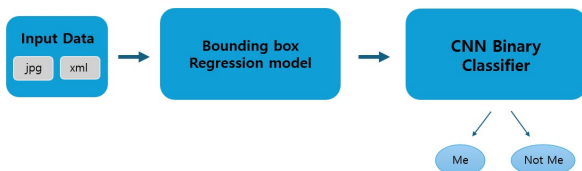


그림 2. Bounding Box Regression 과 CNN 을 사용한 이미지 이진 분류 모델.

먼저, 사용자가 직접 바운딩 박스를 라벨링한 xml 파일과 그에 상응하는 이미지 파일이 입력 데이터로 들어간다. 그 다음 바운딩 박스 회귀 모델에서 데이터의 예측 좌표 값과 실제 좌표 값의 차이를 최소화하는

방향으로 학습 후 바운딩 박스를 예측한다. 바운딩 박스 회귀 모델로 얻은 좌표 값을 입력 이미지와 함께 이진 분류 모델에 제공한다. 그리고 해당 영역의 이미지를 기반으로 나인지 아닌지 판별하는 이진 분류하는 모델로 학습한다. 모델의 성능의 일반화와 높은 정확도를 위해 k-fold 를 이용하여 모델을 훈련하고 평가한다. 값과 실제 좌표 값의 차이를 최소화하는 방향으로 바운딩 박스를 예측한다.

3.3 실행 결과

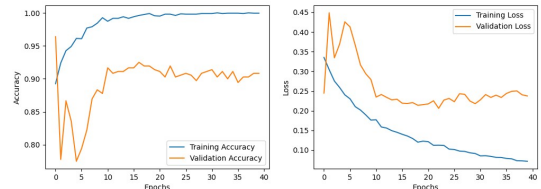


그림 3. Accuracy 와 Loss 로 나타낸 모델 실행 결과

그림 3 에 나와 있듯이 훈련세트는 파란색, 검증세트는 주황색으로 정확도와 Loss 를 표현하였다. 에포크 초반에는 불안정하지만, 에포크가 높을 수록 안정적으로 정확도가 상승하고 loss 값이 하락하는 것을 볼 수 있다.

IV. 결론

CNN 만 사용해 이미지를 분류하는 것에는 한계가 존재한다. 따라서 CNN 으로 구성된 사진 속 내 모습을 판별하는 모델의 정확도를 높이기 위해 Bounding Box Regression 기법을 활용했다.

이 과정을 통해 학습이 필요한 부분만을 학습하여 테스트셋의 내 얼굴을 더욱 효과적으로 검출할 수 있고, 위치 정보도 고려할 수 있게 되어 불필요한 오검출을 방지한다. 이에따라 더 정확하게 분류할 수 있다.

이는 앞으로 복잡한 상황에서 정확한 객체를 탐색하는데 도움을 줄 것이다. 그리고 다양한 환경에서의 적용 가능성을 확대하여 더 넓은 응용 분야 영역에서 활용이 가능할 것으로 기대된다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로

한국연구재단의 지원을 받아 수행된 연구임(과제번호-2022R1A2C1009951).

참 고 문 헌

- [1] Y. LeCun, et al, "Gradient-based learning applied to document recognition," Proc. of the IEEE, 1998.
- [2] Bertozzi, Massimo, et al. "Stereo vision-based vehicle detection." Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511). IEEE, 2000.
- [3] Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." Ijcai. Vol. 14. No. 2. 1995.