

2022 전기 졸업과제 착수 보고서

제조영상 불량 검출 및 분류 인공지능 학습모델 개발

스팀

조장 서지훈

조원 김무영

2022.05

목차

1. 과제 배경 및 목표	3
1.1 과제 배경	3
1.2 과제 목표	3
2. 과제 내용.....	4
2.1 데이터	4
2.2 학습 방안	4
2.3 개발 환경	5
2.3.1 개발 언어 및 IDE	5
2.3.2 관련 라이브러리	5
2.4 시나리오	6
2.5 예상 문제점	6
3. 개발 일정 및 역할 분담	7
3.1 개발 일정	7
3.2 역할 분담	7

1. 과제 배경 및 목표

1.1 과제 배경

현대 사회에서 공장은 기계로 매우 많은 제품들을 제조하고 있다. 기계는 이러한 제품들을 인건비보다 저렴한 가격으로 양산할 수 있다. 하지만 불량품이 발견되지 못하고 출시되는 경우, 기업의 이미지가 크게 훼손되는 등의 손실이 발생할 수 있다. 이러한 손실을 막기 위하여 제조 공정은 생산 과정과 출하 과정에서 이를 주시하지만, 인간의 눈으로 보는 검사에는 한계가 있다. 이를 해결하기 위하여 인공지능 기반으로 불량품을 검출하여 분류하고자 한다. 우리는 현대의 가장 중요한 건축 자재인 철강의 표면 결함 데이터를 바탕으로 불량품을 검출할 것이다.

따라서 이 졸업과제에서는 철강 결함 검출 모델을 개발한다.

1.2 과제 목표

이 졸업과제의 목표는 철강 결함 검출 모델의 개발이므로, 완성된 모델은 다음과 같은 기능을 수행할 수 있어야 한다.

- 데이터는 철강 표면 결함에 따라 전처리가 수행되어야 한다.
- 개발할 모델은 전처리가 수행된 데이터를 바탕으로 불량 철강 분류를 위한 지도 학습(supervised learning)을 수행해야 한다.
- 또한, 지도 학습이 완료된 후의 모델은 학습되지 않은 철강 데이터의 표면의 흠집을 판단하여 정상적인 철강과 불량 철강을 분류할 수 있어야 한다.

2. 과제 내용

2.1 데이터

데이터는 www.kaggle.com 을 참조하였다.

- [1. 철강 제조에서 발견되는 결함의 위치와 유형을 바탕으로 결함품을 찾고 분류하는 데이터](#)
- [2. 고온에서 압연한 강철의 대표적인 표면 결함 6 가지의 데이터](#)

위와 같은 두 데이터를 이용한다.

1 번 데이터는 [1, 2, 3, 4]의 네 종류가 존재하는 'ClassId'라는 레이블로 결함을 분류한다. 결함이 존재하는 이미지 파일과, 관련 csv 파일이 제공된다. csv 파일은 ClassId 에 따라 labeling 되어있다.

2 번 데이터는 6 가지의 표면 결함(rolled-in scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In) and scratches (Sc))이 존재한다. 결함이 존재하는 이미지 파일과, 관련 xml 파일이 주어진다.

xml 파일은 결함의 종류에 따라 labeling 되어있다.

위 데이터들에 대해 전처리를 수행하여 학습 알고리즘에 사용한다.

2.2 학습 방안

우리는 결함을 발견하는 것이 목표이므로, 철강의 표면에서 결함이라는 특징만을 발견하면 된다.

따라서 학습 알고리즘으로 이미지의 특징을 추출하여 class 를 분류하는 CNN(Convolutional Neural Network)을 사용할 것이다.

예를 들어, 1 번 데이터의 ClassId 레이블을, 2 번 데이터의 6 가지 결함을 특징으로 학습시킨다.

2.3 개발 환경

2.3.1 개발언어 및 IDE

인공지능 모델을 만들어야 하므로 언어는 Python 3.7 을 채택한다.

인공신경망을 활용하므로 GPU 를 사용하는 것이 좋으며, 결과값 분석이 필요하므로 IDE 로는 jupyter notebook 을 주로 사용할 것이다.

2.3.2 관련 라이브러리

1. 데이터

1 번 데이터의 전처리를 위해 csv 파일의 처리가 필요하다.

이를 위하여 Python 외부 라이브러리 pandas 를 사용한다.

pandas 를 통하여 csv 파일을 데이터프레임으로 변환한다.

데이터 프레임의 관리를 위하여 행렬 등의 연산을 지원하는 외부 라이브러리 numpy 를 사용한다.

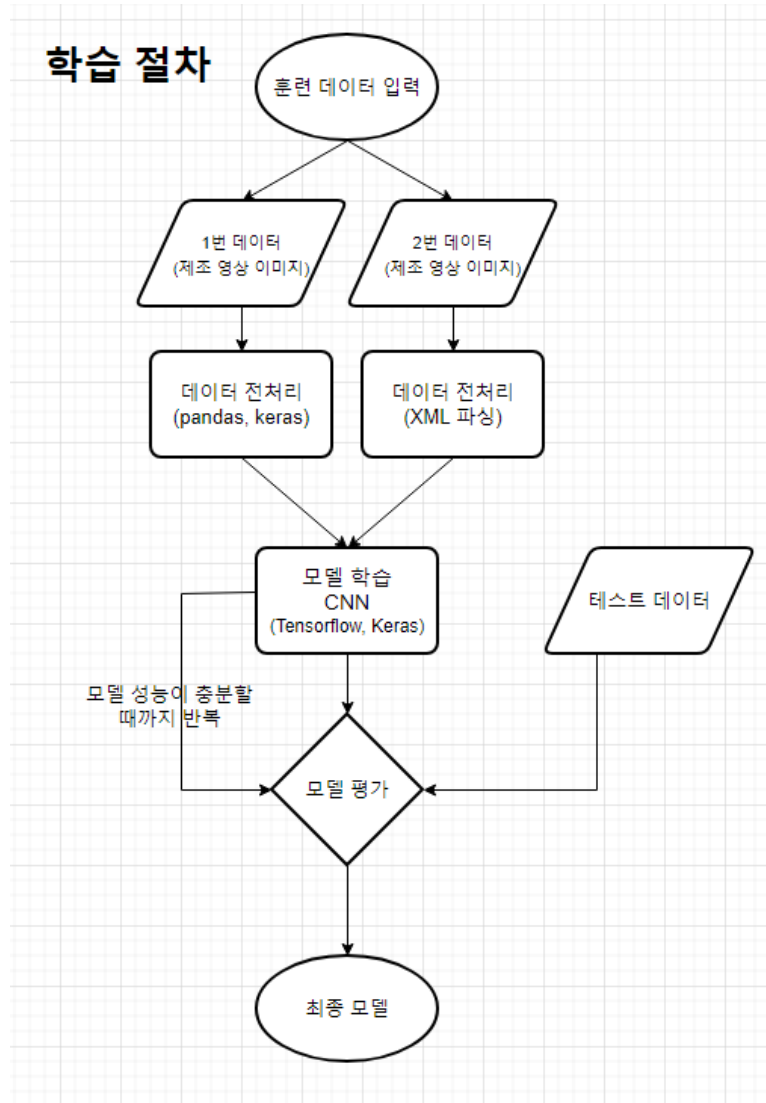
2 번 데이터의 전처리를 위하여 XML 파싱이 필요하다.

따라서 Python 2.5 부터 내부 라이브러리에 포함되어 있는 ElementTree 를 이용하여 파싱한다.

2. 모델 개발

오픈소스 머신러닝 라이브러리인 TensorFlow 와 TensorFlow 위에서 작동하며 사용하기 쉬운 라이브러리인 Keras 를 사용할 것이다.

2.4 시나리오



2.5 예상 문제점

CNN은 간단한 기계 학습 알고리즘에 비하여 속도가 매우 느린 편이다. 철강의 생산과 출하 과정까지 결함 검출 모델의 작업이 수행되지 못한다면, 출하된 철강을 다시 회수해야 하는 등 처리 속도에서 문제가 생길 수 있다. 모델의 크기가 지나치게 크거나 충분한 컴퓨터 자원이 확보되지 못한 상황에서 모델을 사용해야 한다면, 학습 알고리즘의 수정이 필요할 수 있다.

3. 개발 일정 및 역할 분담

3.1 개발 일정

5월 3주 ~ 6월 4주	기본 모델 작성
7월 1주 ~ 7월 2주	모델 최적화
7월 3주 ~ 7월 4주	중간 보고서 작성
8월 1주 ~ 9월 2주	전체 시스템 검증 및 수정
9월 3주 ~ 9월 4주	최종 보고서 작성

3.2 역할 분담

서지훈	(공통)보고서 작성 철강 결함 검출 모델 초기 형태 개발 모델 학습 파라미터 값 조정 완성된 모델 최종 검증
김무영	(공통)보고서 작성 훈련용 데이터 전처리 모델의 속도 및 사용 메모리 최적화 모델 결과 분석 및 재 학습 판단