

# 병해충 AI 학습용 데이터 구축 가이드 (Draft)

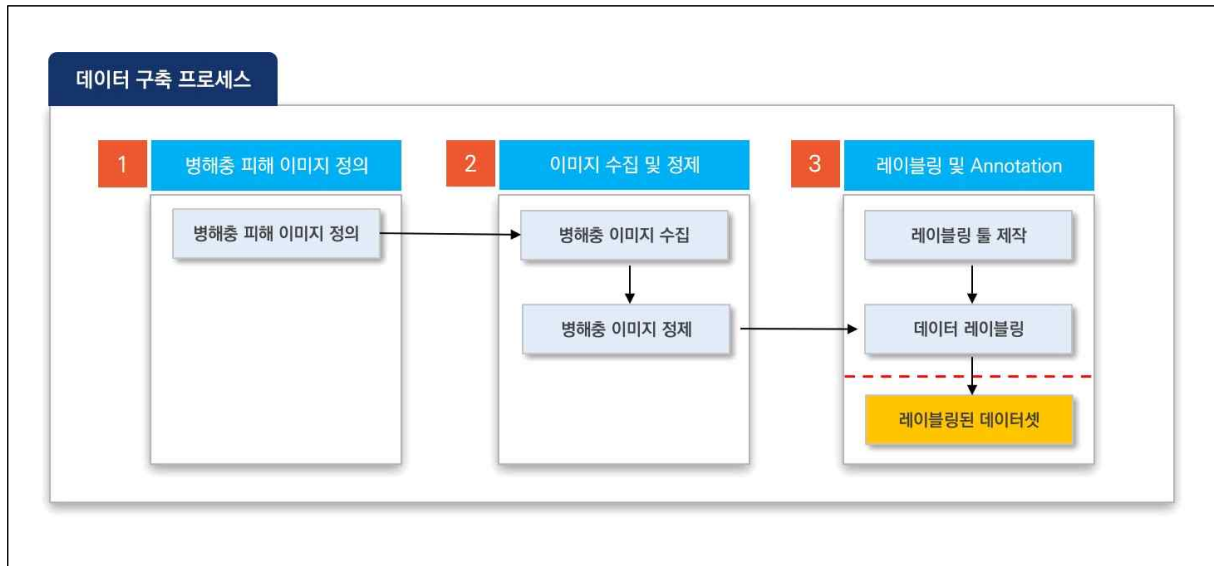
2021.05.03.

농촌진흥청 과수화상병문제병해충연구단

## 목 차

<b>I. 데이터 구축 개요</b> .....	<b>1</b>
1.1 병해충 피해 이미지 정의 .....	1
1.2 병해충 피해 이미지 수집 및 정제 .....	1
1.3 병해충 메타 정보 구축 .....	1
1.4 Labeling 및 Annotation .....	1
<b>II. 병해충 메타 정보 구축</b> .....	<b>2</b>
<b>III. 병해충 이미지 촬영 가이드</b> .....	<b>3</b>
<b>IV. 데이터 Labeling 및 Annotation</b> .....	<b>8</b>
1. Annotation 도구 사용 프로세스 .....	9
2. 올바른 Annotation 예시 .....	17

## 1. 데이터 구축 개요



### 1.1. 병해충 피해 이미지 정의

- 스마트폰으로 촬영된 영상(이미지) 자료에서 병해충을 구체적으로 정의하는 단계

### 1.2. 병해충 피해 이미지 수집 및 정제

- 정의한 병해충을 포함한 이미지 자료를 수집 및 정제하는 단계
- 병해충 이미지 데이터는 기존에 농촌진흥청에서 수집된 자료를 사용
- 또한 인터넷에 오픈되어 있는 이미지 데이터 셋도 함께 활용
- 인터넷에서 수집한 이미지에는 불필요한 잡음이 포함되어 있을 수 있기 때문에 검수 시스템을 통해 최종적으로 양질의 데이터 셋을 선별하는 것을 목표로 함

### 1.3. 병해충 메타 정보 구축

- 병해충에 대한 메타 정보를 구축하는 단계
- 메타 정보에는 촬영된 작물의 작물명, 발병시기, 발병환경, 피해증상, 방제방법, 적용약제 등 정보로 구성됨

### 1.4. Labeling 및 Annotation

- 인공지능 모델 학습을 위한 데이터 레이블링 및 Annotation 단계
- 병해충 이미지의 레이블(예: 화상병, 검은별무늬병)과 이미지내 병징 부위의 정보를 저장하는 과정

## 2. 병해충 메타 정보 구축

- 병해충 정보를 효율적으로 관리 및 활용하기 위해 병해충 메타 정보를 별도로 구축
- 병해충 메타 정보

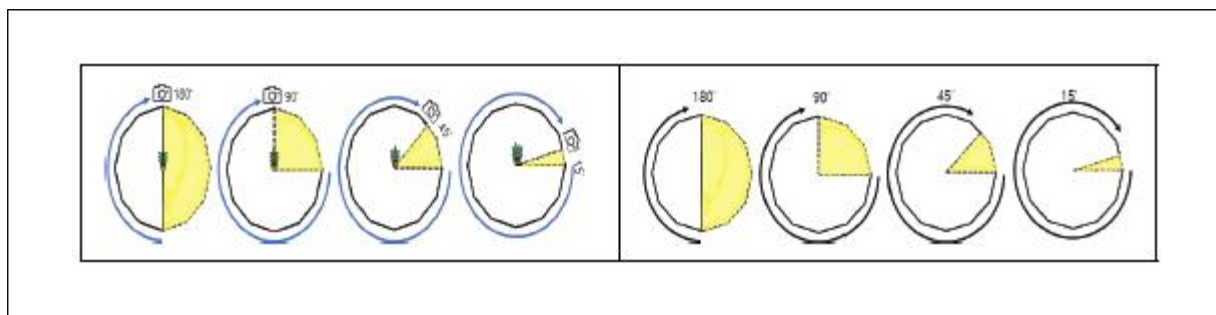
구분	데이터 타입	설명
농작물 명	String	• 농작물 카테고리(예: 사과, 배 등)
병해충 구분	String	• 병 혹은 해충 구분
병해충 명	String	• 병해충 명(예: 화상병, 검은별무늬병)
발병 시기	String	• 발병 시기 기재(예: 4월~7월)
발병 작물	String	• 발병 가능한 농작물
발생환경	String	• 발생환경 기술
피해증상	String	• 피해증상 기재
방제방법	String	• 방제방법 및 적용 가능한 약제 정보 기재

- 병해충 메타 정보 json 예시

```
{  
  plant_name: "배", ## 작물명  
  disease_or_pest: "병", ## 병해충 구분  
  disease_name: "화상병", ## 병해충 명  
  duration: "4월~7월", ## 발병 시기  
  occurred_plant: "배, 사과", ## 발병 작물  
  environment: "...", ## 발생환경  
  symptom: "...", ## 피해증상  
  prescription: "..." ## 방제방법 및 약제  
}
```

### 3. 병해충 이미지 촬영 가이드

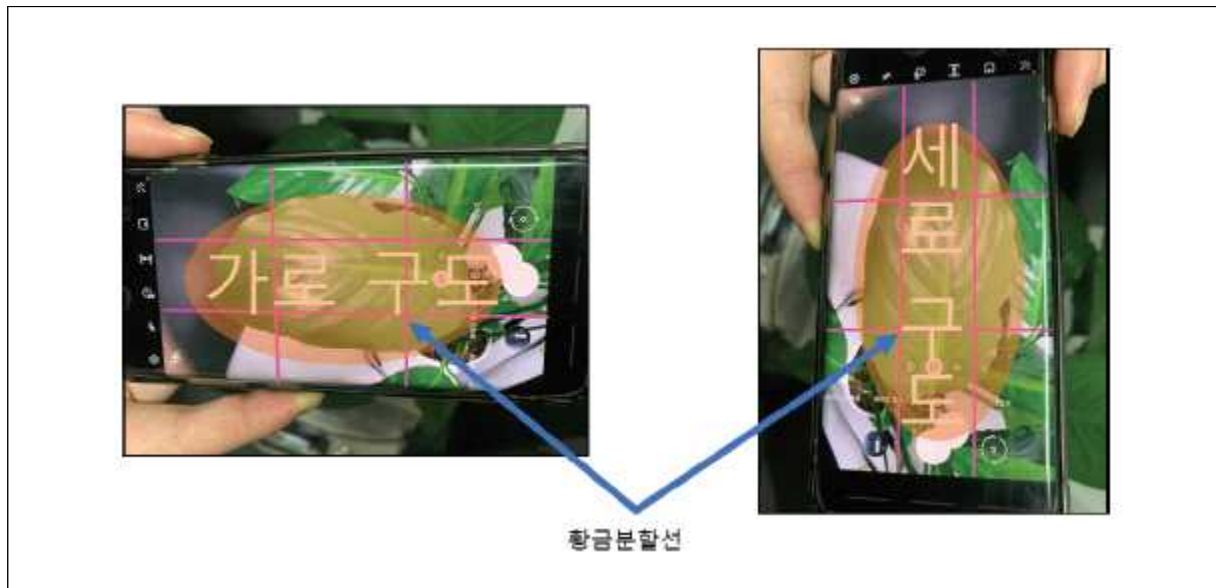
- 촬영 장비: 스마트폰
- 일관된 이미지 품질 및 수량 범위를 위해 촬영 장비, 높이, 각도, 거리, 조명 등 기준 명시
- 위치정보(GPS) 활성화
- 플래시를 사용한 촬영 금지 (카메라에서 플래시 없음 설정) 필요시 조명 설치 권고
- 병해충 이미지를 촬영시 최대한 근접한 거리, 병징이 선명하게 보이도록 함
- 촬영 데이터는 기본적으로 해상도 만족을 위해 Full HD 장축 1920(Pixel)픽셀 이상 확보
- 파일 포맷 : JPG, PNG, TIFF 형식(압축 형식을 사용하므로 손실을 확인 필요)
- 장비별, 객체별, 시간별, 종류별, 사람별, 지역 별 검토
- 데이터 획득의 정확도를 높이기 위해 이미지 획득 시 촬영 각도 참조기준에 따라 촬영
  - 촬영 환경에 따라 촬영 각도 외 특정 각도 가능



〈촬영 각도 예시〉

- 촬영 데이터 각 건당 최소 8장 이상 촬영 : 원거리 1장 필수, 정면, 위, 아래, 각 15도, 45도, 90도, 180도 등 각도를 달리한 이미지 데이터
- 영상(동적/정적) 이미지 화면 구성
  - 화면비율: HD가 디지털 영상의 표준으로 자리 잡았기 때문에 4:3보다는 16:9 화면이 일반적임
  - 화면을 좌, 우 또는 위, 아래로 나누었을 때 어느 쪽에 피사체를 위치시키는지에 따라서 균형이 유지되거나 깨지기 때문에 균형 있는 이미지 획득 필요
  - 사진이 규정된 촬영 조건과 촬영 방식을 준수하였는지 확인
- 황금비율 영상(동적/정적) 이미지 사진(예시)





〈황금비율 촬영 기법 예시〉

○ 획득 단계부터 식별해서 병충해(병명, 충 명 등) 이미지 획득 필요



〈정상작물과 해충작물 피해 증상 데이터 획득〉



〈생육기별 형태 (좌: 유충, 우: 성충) 영상(동적/정적) 이미지 획득

- 촬영 시 Zoom-In, Zoom-Out에서 객체의 화면이 ‘황금비율의 예시’와 동일하게 해상도와 색상, 화소 등이 최대한 원판과 흡사하게 표현되어야 하며, 객체의 상이 틀어지거나 변경되지 않도록 확인 필요.(전체 화면의 70% 이상 식별이 용이하도록 채움)
- 정상작물을 시작으로 초기, 중기, 말기로 해충 피해 구분별 영상(동적/정적) 이미지 획득
- 작물 부위별, 감염 정도별, 시기별 다양한 피해 증상에 대한 데이터 확보
- 노지 작물별 발생하는 해충의 생육기별 형태 및 피해 증상 영상 데이터 획득
- 조사대상, 해충 종류, 계절의 환경적 요인, 생육단계, 획득지역, 획득 시간 등 정확한 식별 가능한 수준의 데이터 확보 필요
- 다양한 형태의 학습데이터를 획득하기 위해 거리, 각도, 해상도, 조도 등 활용
- 획득 장비: 스마트폰 단말기/스마트폰 단말기(전용App설치)/디지털 카메라 장비 활용
- 데이터 구축 목적에 따라 분야별 메타데이터를 정의하고 이러한 기준에 맞춰 메타데이터와 함께 획득된 데이터는 정제 전 데이터베이스(DB)에 저장함



**〈올바른 병해충 이미지 촬영 예시〉**

- 여러 개의 병이 함께 있는 경우 촬영 대상에서 제외
- 병해충 이미지에서 병징이 양면으로 발생시 각 면의 병징을 별도로 촬영



병징이 양면으로 있을 경우



○ 한 장의 이미지를 여러 각도로 촬영

다양한 각도로 촬영





○ 잘못된 병해충 이미지 촬영은 아래와 같음

올바르지 않은 결과물 예시	사유
	<ul style="list-style-type: none"> <li>• 촬영 거리가 너무 멀어서 병해충 병징이 뚜렷하게 나타나지 않음</li> </ul>
 <p>photo 2-20 - K. D. Hickey</p>	<ul style="list-style-type: none"> <li>• 지나친 확대에 의한 해상도 저하</li> </ul>
	<ul style="list-style-type: none"> <li>• 높은 밝기로 인한 해상도 저하</li> </ul>

## 4. 데이터 Labeling 및 Annotation

- 데이터 Labeling 시 식별 분류는 인공지능 학습의 품질을 결정하는 핵심 업무로서 특성을 명확하게 규정해 주는 것이 지도 학습에서의 결과물로서의 가치를 높이는 기준이 됨. 이를 위한 고려사항으로 획득 가능성과 활용데이터의 보안 문제, 데이터 정확성 등을 확인하고 고민해서 라벨링 단계를 진행해야 함
- 식별된 특성을 통해 Labeling에 필요한 Labeling 기준 및 Annotation 속성항목 작성
- 전처리 진행 중에 추가사항 지속적 업데이트
- Labeling 후 데이터 시험과 성능 등 정확성과 적합성 검증을 위한 데이터셋의 비율이 필요하며, 인공지능 개발자와의 협의를 통해 지정
  - ※ 예) 학습데이터(80%), 검증데이터(10%), 시험데이터(10%)
- 일반적인 상식으로 수집된 데이터에서 약 1,000장 정도의 이미지를 가공하여 모델링에 접목시킨 후 모델과 적합성을 확인함
- 인공지능 개발에서 일반적으로 사용되는 Labeling 용어 및 분류체계를 준수함
- 모든 산출 데이터는 표준화된 포맷을 지원할 수 있도록 CSV, JSON 등으로 묶어서 제공 필요
- Annotation 작업 방식: **Bounding Box**
  - 객체를 직사각형 모양의 박스 안에 포함되도록 그리는 라벨링 방법으로 데이터 라벨링 작업에서 가장 일반적으로 사용
  - Bounding Box는 객체를 전체가 커버되도록 하며, 박스 안에 객체 이외의 여백을 최소화 하도록 지정

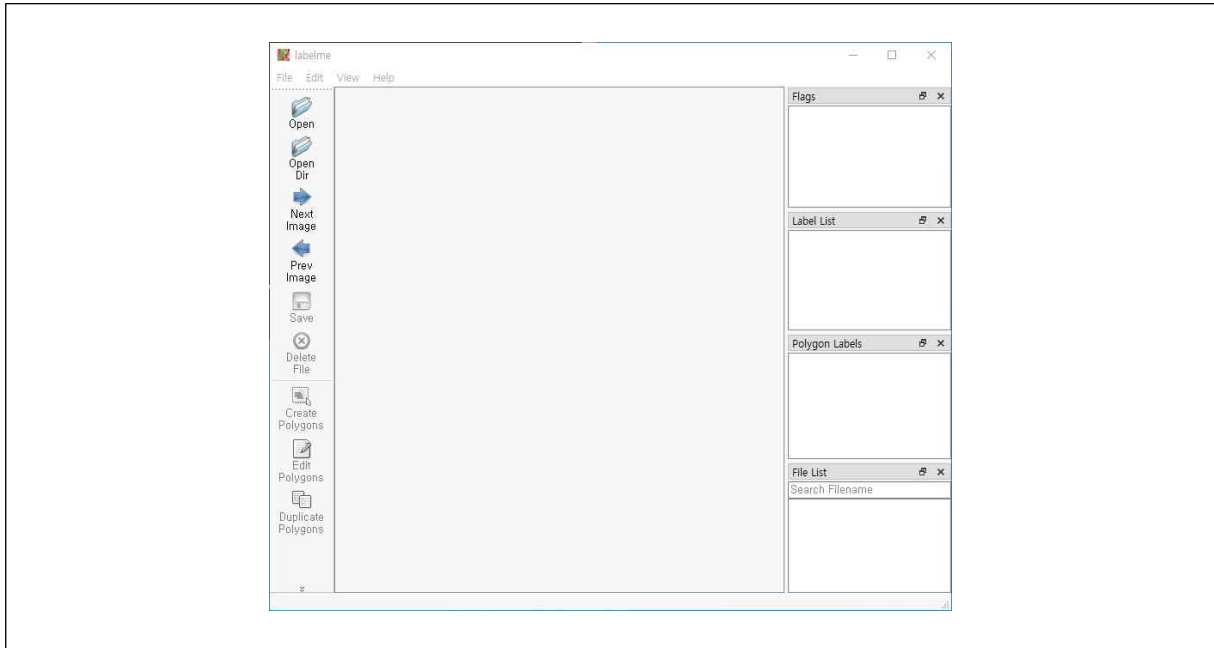


〈Bounding Box 예시〉

## 4.1. Annotation 도구 사용 프로세스

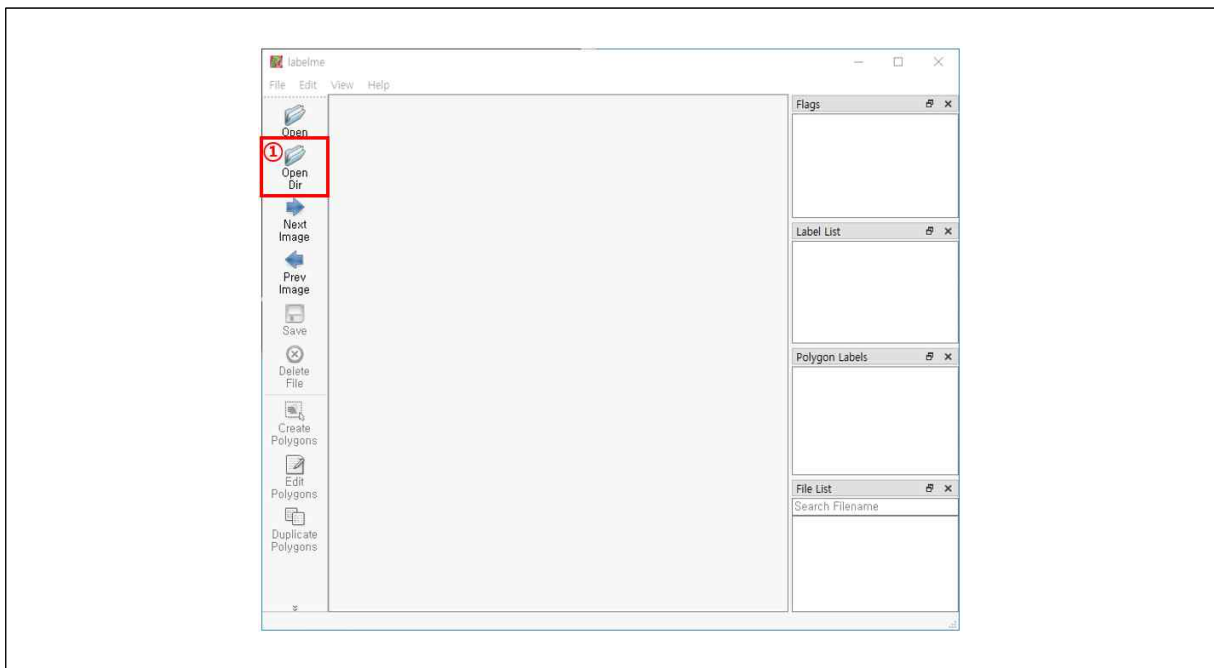
### ① Labeling도구 실행

- labeling 도구를 실행한 화면은 아래와 같음

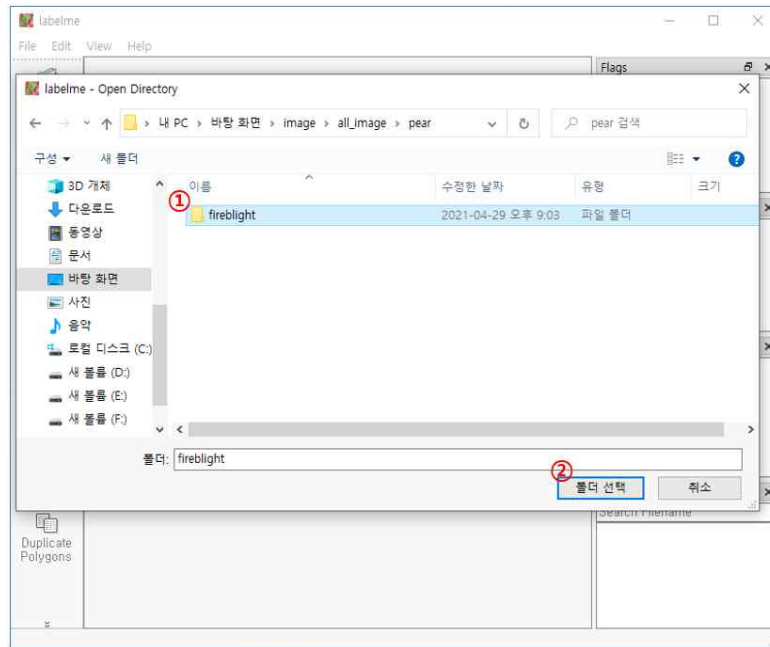


### ② Labeling을 진행할 이미지가 위치한 폴더 선택

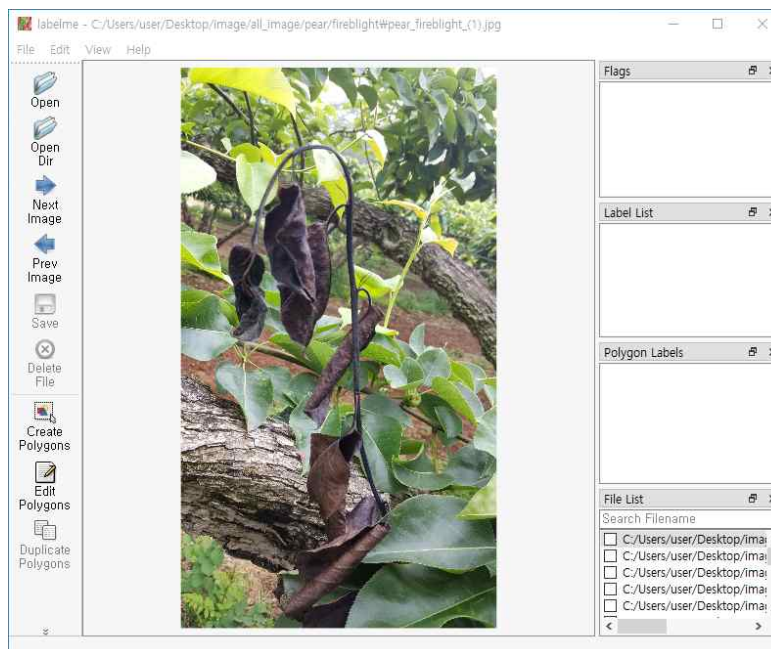
- “Open Dir” icon을 클릭해 labeling을 진행할 이미지가 저장된 폴더를 선택함



- “Open Dir” icon을 클릭하면 컴퓨터에 있는 폴더를 선택할 수 있는 창이 나타나고 창을 통해 labeling을 진행하려는 이미지가 포함된 폴더를 선택한 후 “폴더 선택” 버튼을 클릭함



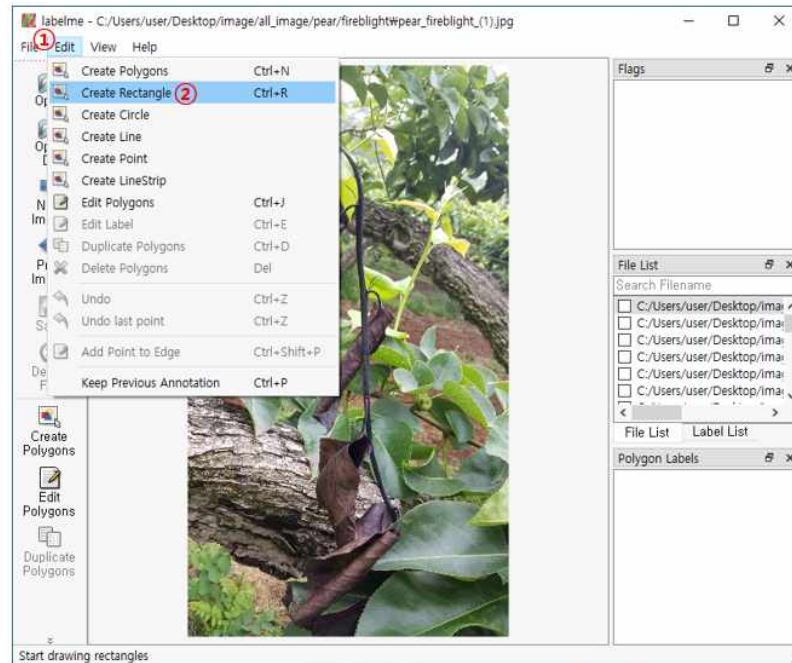
- “폴더 선택” 버튼을 클릭하면 폴더에 있는 첫 번째 이미지가 프로그램 main panel에 나타나고 폴더에 있는 전체 이미지는 오른쪽 아래 위치에 있는 “File List” panel를 통해 확인이 가능함



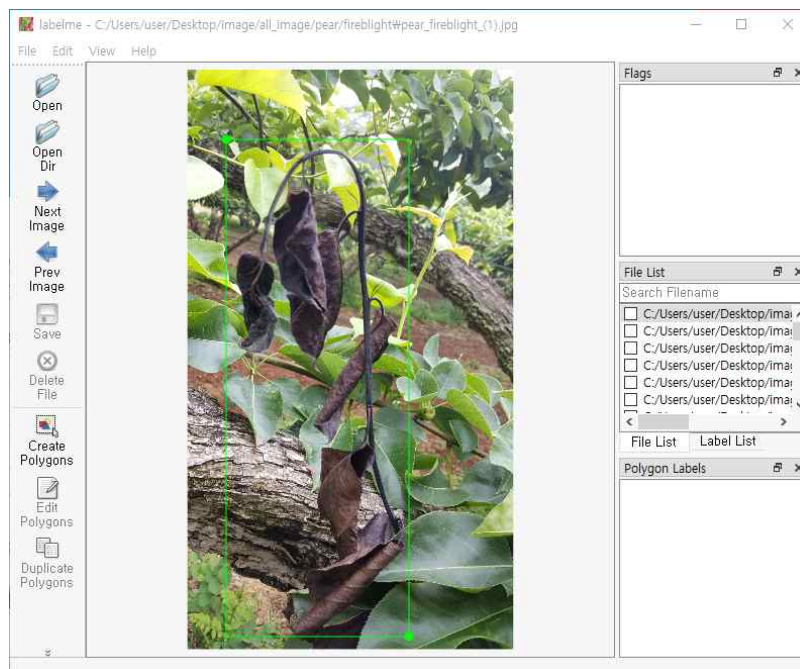


### ③ 선택된 이미지에 대해 labeling 작업 진행

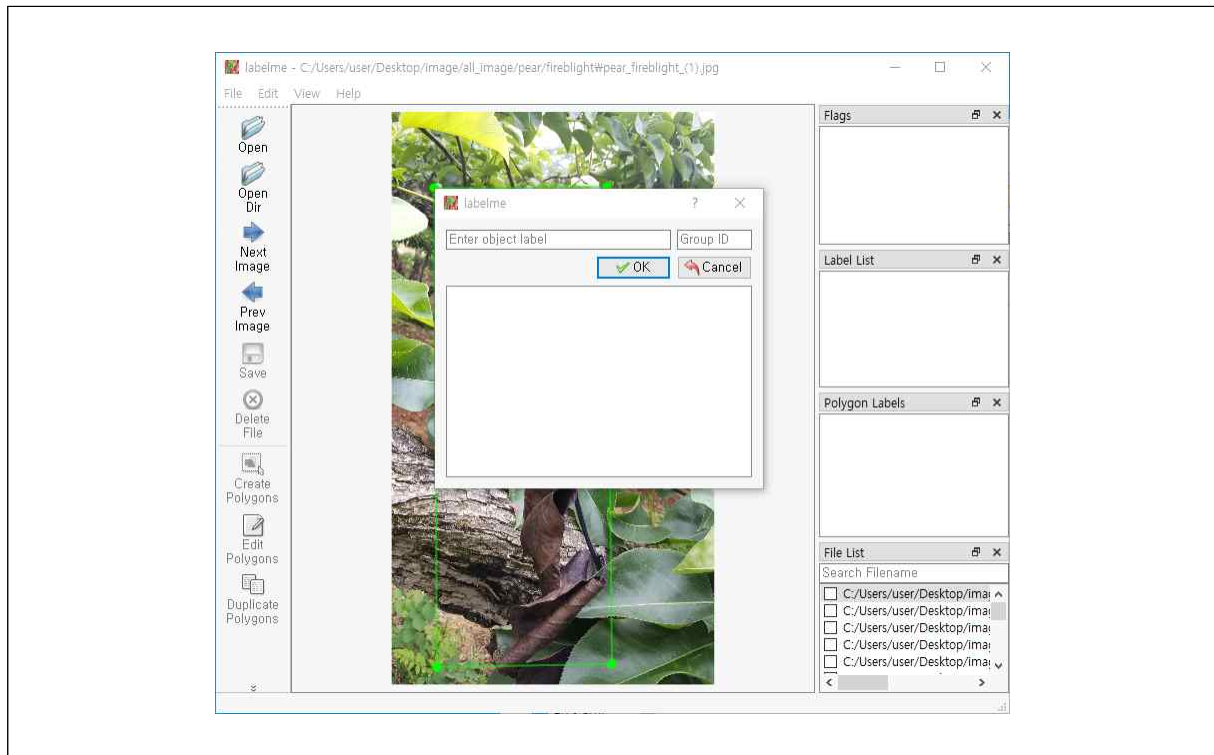
- 화면 상단에 있는 “Edit” 버튼을 클릭한 후 “Create Rectangle”를 클릭해 현재 보여주고 있는 이미지에 대해 labeling 작업을 진행함



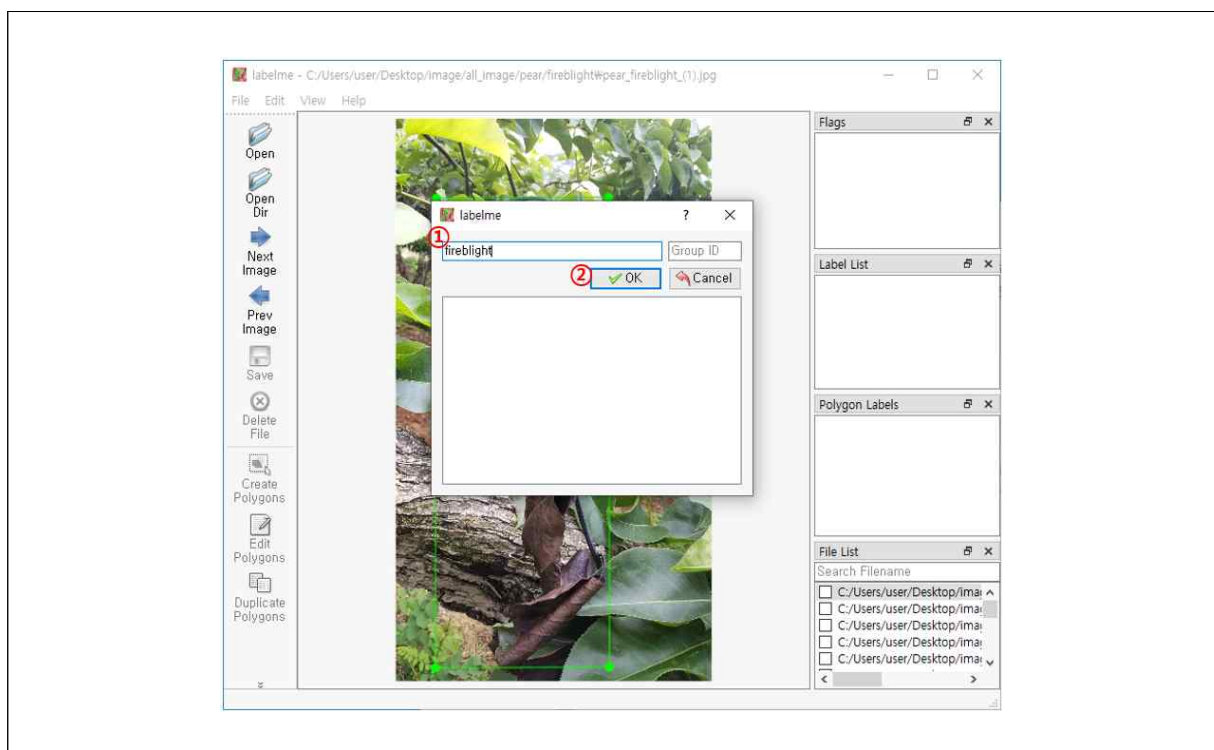
- “Rectangle”을 사용해 정확한 병징의 위치를 사각형(bounding box)으로 labeling 함



- “Rectangle”을 사용해 bounding box가 완성되면 아래와 같이 label 정보를 입력할 수 있는 화면이 나타나게 됨

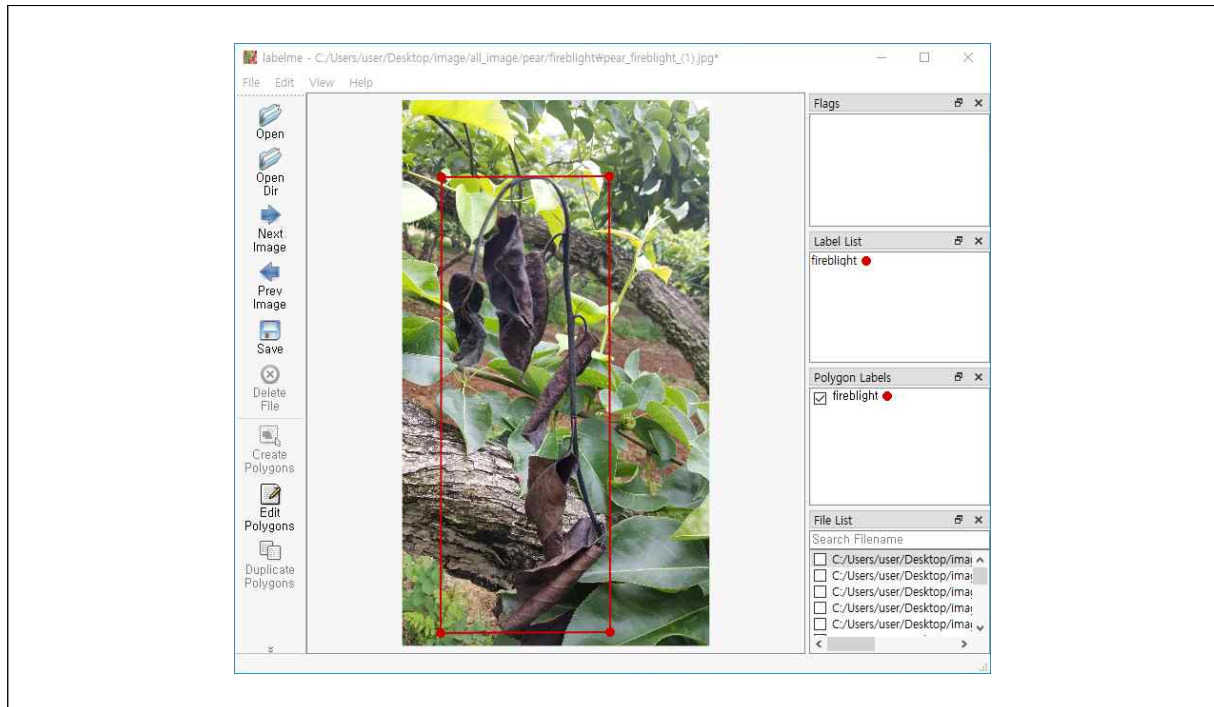


- Label 정보에 병해충 이름을 입력하고 “OK” 버튼을 클릭하면 하나의 instance(병징 부위)에 대한 labeling 작업이 완성됨

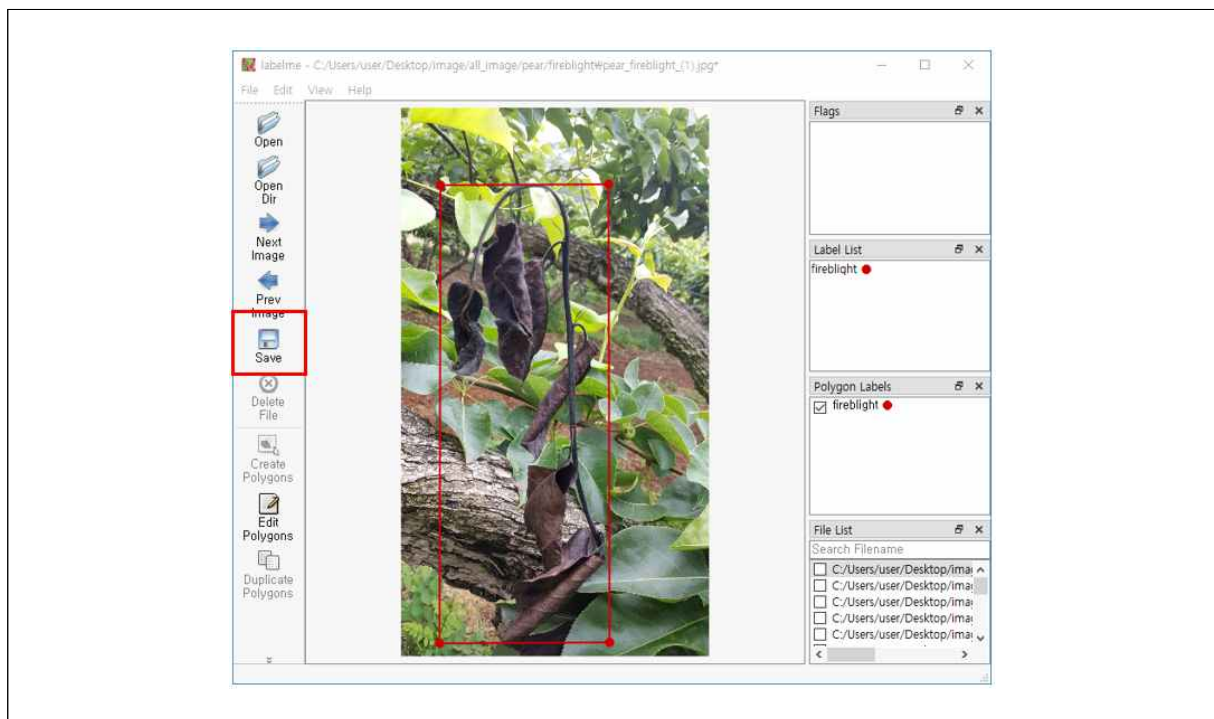


#### ④ Labeling 결과 확인

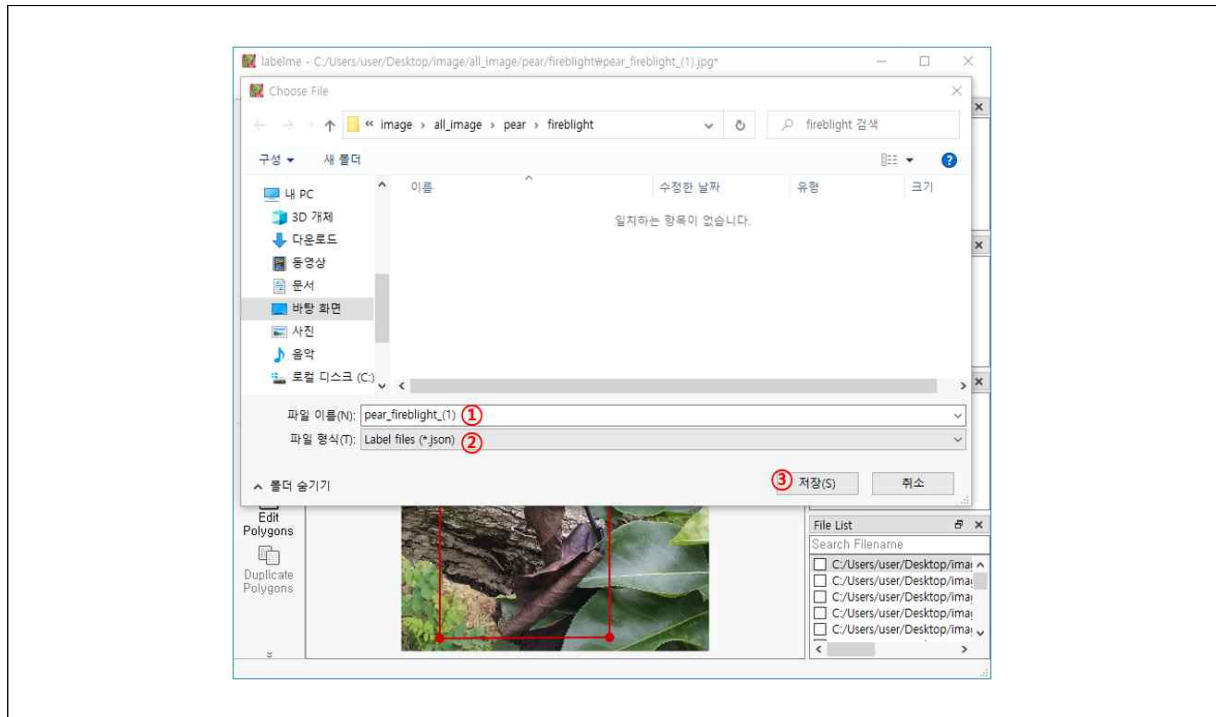
- “OK” 버튼을 클릭해 하나의 instance에 대한 labeling 작업을 완성하면 아래와 같이 labeling 결과를 확인할 수 있음



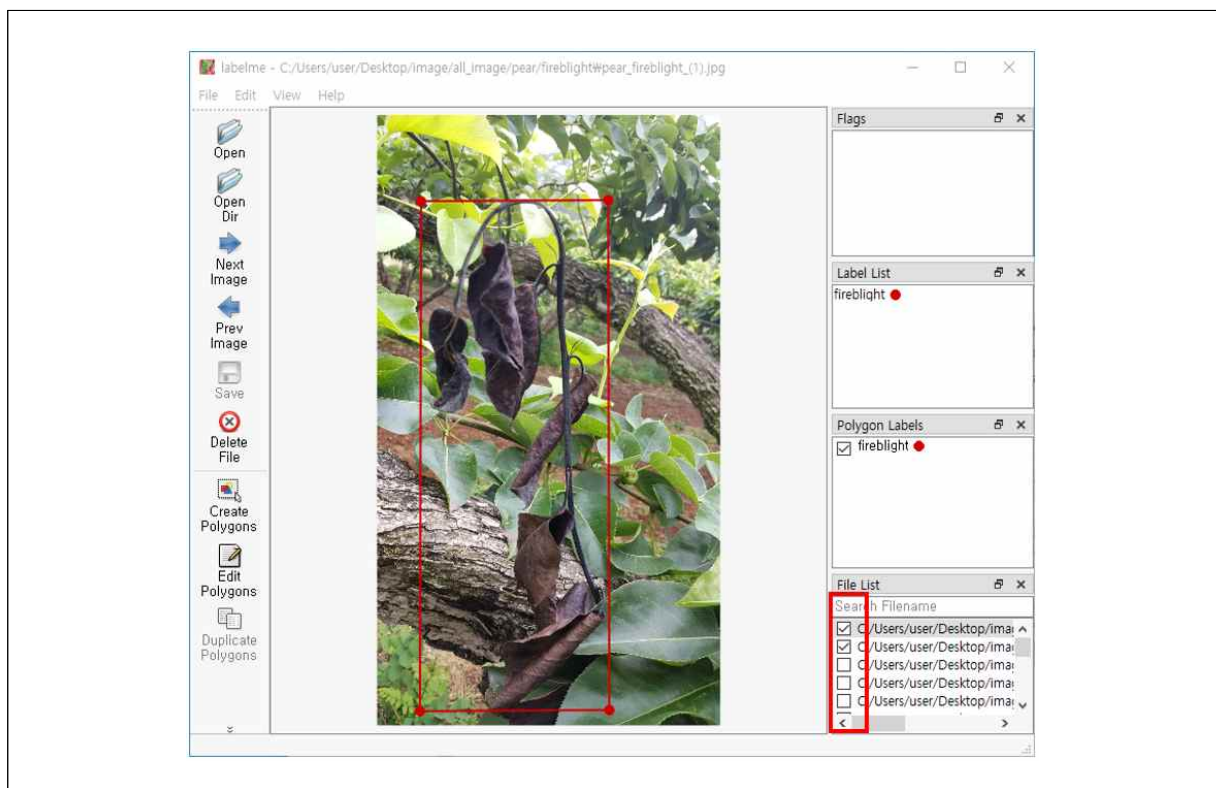
- 이미지에 하나 이상의 instance가 존재하면 ③번 작업을 반복해 진행함
- 해당 이미지에 대한 labeling 작업이 완료되었으면 “Save” 버튼을 클릭해 작업한 이미지에 대한 labeling 정보를 저장함



- “Save” 버튼을 클릭하면 label 파일을 저장하는 창이 나타나는데 파일 이름을 이미지의 이름과 동일하게 설정하고 파일 형식을 Label files (\*.json)으로 설정한 후 “저장” 버튼을 클릭함



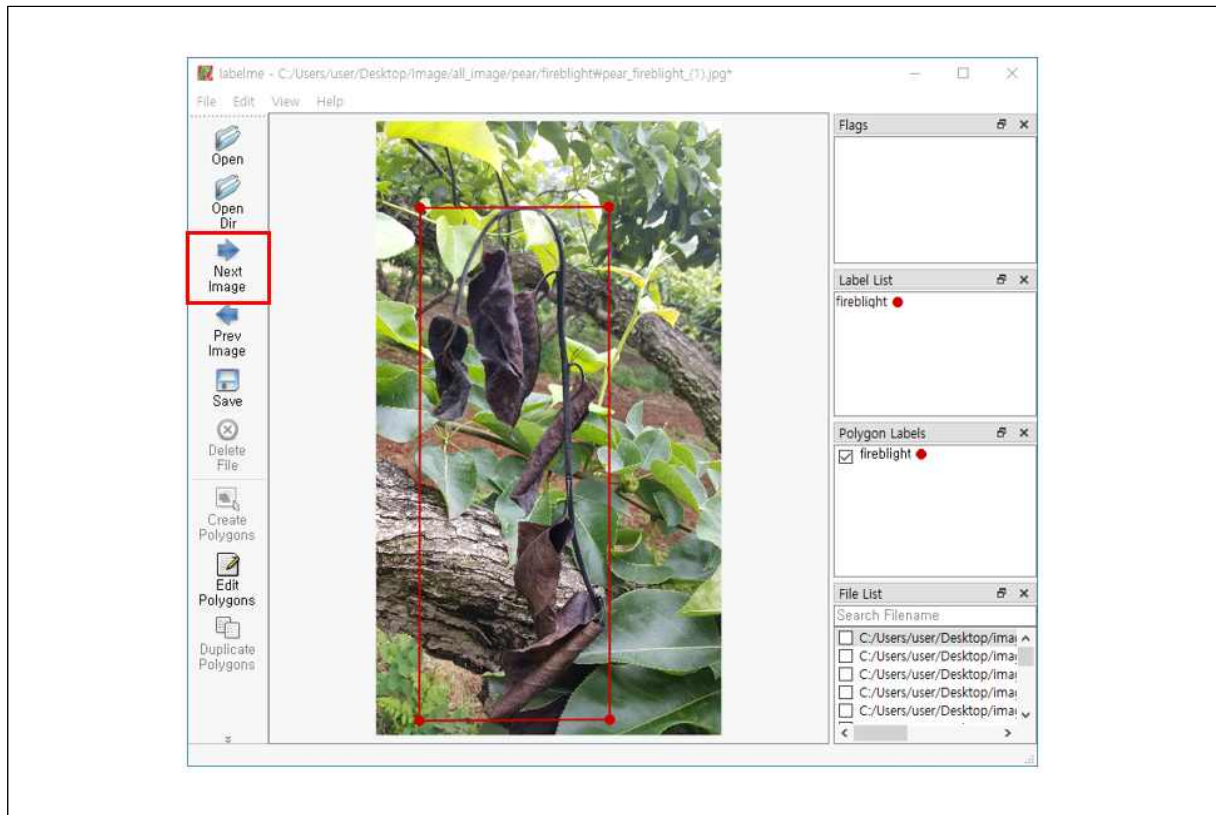
- 저장된 이미지는 오른쪽 아래 “File List” panel의 파일 리스트 앞의 체크 표시를 통해 확인할 수 있음





### ⑤ 다음 이미지에 대해 labeling 작업 진행

- “Next Image” 버튼을 클릭해 다음 이미지에 대해 ③번과 ④번에서 명시한 작업을 진행함



- 폴더에 있는 모든 이미지에 대해 labeling 작업이 완료되면 프로그램을 종료하거나 “Open Dir” 버튼을 클릭해 다른 디렉터리에 있는 이미지에 대해 작업을 진행할 수 있음

### ⑥ 저장된 Annotation 정보 확인

- 한 장의 이미지에 대해 labeling을 진행해 얻은 annotation 파일의 내용은 아래와 같음

```

1  {
2    "version": "4.5.7",
3    "flags": {},
4    "shapes": [
5      {
6        "label": "fireblight",
7        "points": [
8          [
9            136.85714285714283,
10           246.4285714285714
11          ],
12          [
13            783.2857142857142,
14            2003.5714285714284
15          ]
16        ],
17        "group_id": null,
18        "shape_type": "rectangle",
19        "flags": {}
20      }
21    ],
22    "imagePath": "pear_fireblight_(1).jpg",
23    "imageData": "/9j/4AxPuHp+dbNtdrKuQRmuKrDl3nBrJtbxomwTkGtmJxKgYda64yuhH//2Q==",
24    "imageHeight": 2048,
25    "imageWidth": 1152
26  }

```

- Annotation 정보의 각 fields의 내용은 아래와 같음




구분	설명
version	labeling을 할 때 사용한 labelme 도구의 버전
flags	이미지 분류 문제를 위한 annotation을 진행할 때 사용됨
shapes	리스트 형식으로 bounding box들의 정보를 담고 있음
imagePath	annotation 파일 기준 이미지의 상대 위치(relative path)
imageData	이미지를 Base64로 인코딩해 저장함
imageHeight	이미지의 높이
imageWidth	이미지의 너비

- “shapes” field에는 각 bounding box에 대한 정보가 리스트로 포함되어 있음
- 리스트의 각 원소에는 아래와 같은 bounding box에 대한 정보가 포함되어 있음

구분	설명
labels	Bounding box의 라벨값
points	Rectangle의 왼쪽 위의 정점 및 오른쪽 아래 정점의 x, y좌표
group_id	Bounding box의 group id로 annotation 할 때 설정 가능

구분	설명
shape_type	Annotation을 진행할 때 사용한 방법, rectangle
flags	이미지 분류 문제를 위한 annotation을 진행할 때 사용됨

## 4.2. 올바른 Annotation 예시

	<ul style="list-style-type: none"> <li>Bounding box가 병징을 모두 포함하며 box의 크기가 외각선에서 최대한 타이트하게 Annotation</li> </ul>
	<ul style="list-style-type: none"> <li>사진에서 특정할 수 있는 병징의 외각선에 비해 Bounding box가 매우 크게 그려져 있어 잘못 Annotation 된 데이터로 분류</li> </ul>
	<ul style="list-style-type: none"> <li>사진에서 병징을 특정할 수 있는 외각선 부분보다 Bounding box가 작게 Annotation 되어 있어 적합하지 않음</li> </ul>