# Fast R-CNN을 이용한 Malware Classification

*인공지능 기술을 활용한 대략의 악성코드  분석*

2015004920 임우태
2016025514 서현아

# Preview

- 프로젝트 개요
- 프로젝트 내용
- Fast R-CNN 선정 이유
- 프로젝트 기대 효과
- 프로젝트 추진 계획

# 프로젝트 개요

**기존의 악성코드 탐지 방식**

악성코드의 특별한 패턴을 분석하여  시그니처로 안티바이러스 프로그램에 등록

**기존 방식의 한계점**

최근 악성코드의 변종이 크게 증가하는 추세 -> 기존 방식은 변종이나 제로데이 공격 대응 어려움

**해결책**

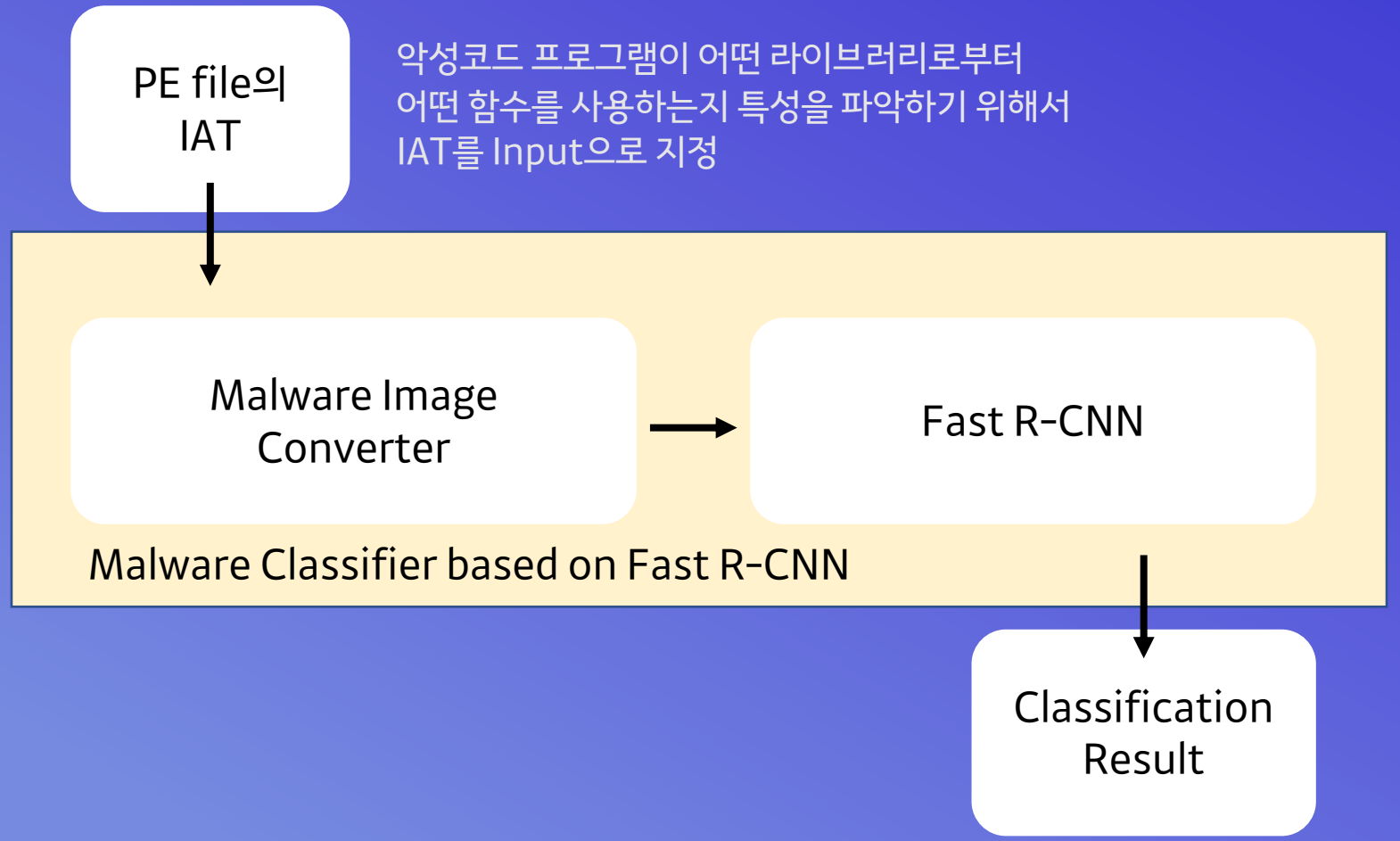
인공지능 (딥러닝, 머신러닝 등)이 기존 방식의 한계점을 해결하기 위해 악성코드 탐지에 적용됨

**인공지능을 이용한 연구 동향**

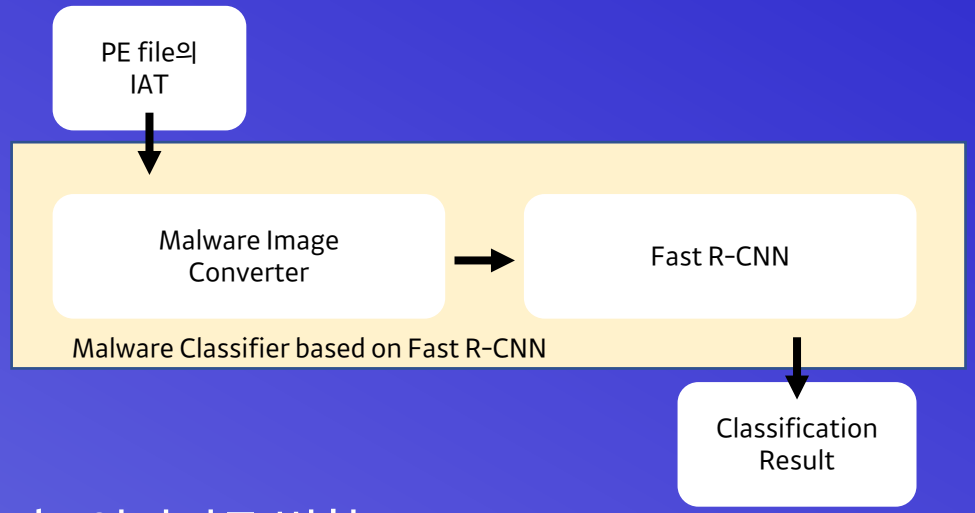
악성코드 파일로부터 특징을 추출 (예 파일을 이미지화)

추출한 특징을 갖고 딥러닝 모델을 트레이닝

# 프로젝트 내용


PE file의
IAT
악성코드 프로그램이 어떤 라이브러리로부터
어떤 함수를 사용하는지 특성을 파악하기 위해서
IAT를 Input으로 지정
Malware Image
Converter
Fast R-CNN
Malware Classifier based on Fast R-CNN
Classification
Result

# 프로젝트 내용




## 악성 여부를 labeling하는 모델을 학습

### Malware Image Converter

API의 sequential 특징이 담긴 PE 파일의 ***IAT***를 2D gray scale 이미지로 변환

### Fast R-CNN을 이용한 malware classify

2D gray scale 이미지를 input으로 fast R-CNN을 사용하여

Malware classification model 학습

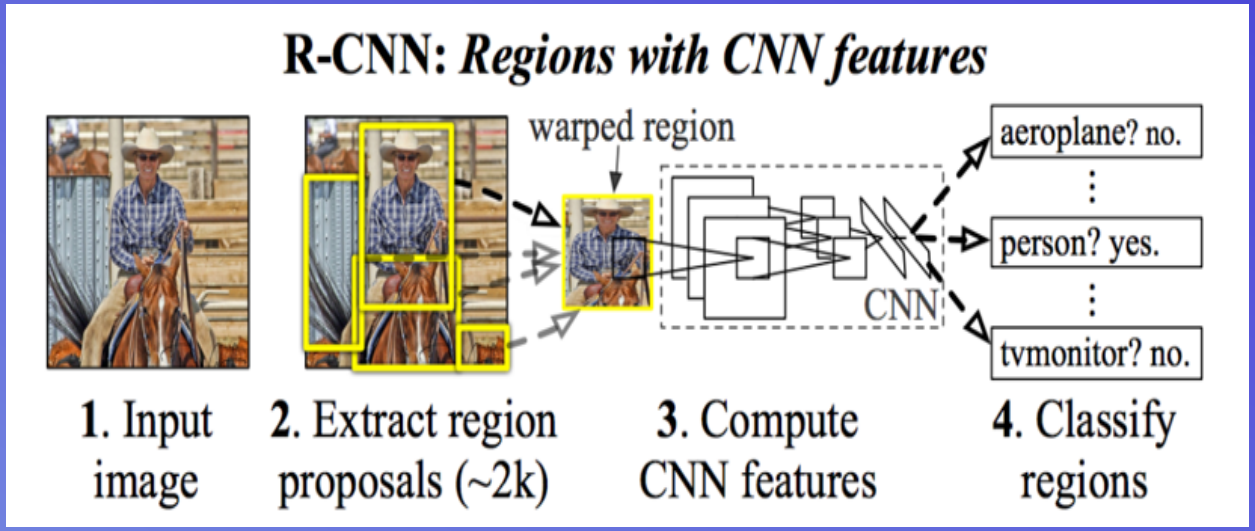
# Fast R-CNN 선정 이유

## R-CNN의 한계

1) 평균적으로 이미지 하나 당 2000번의 forwarding

-> 연산 횟수 증가 -> ***training speed*** 측면에서 한계

2) 세 개의 model을 training -> ***performance*** 측면에서 한계

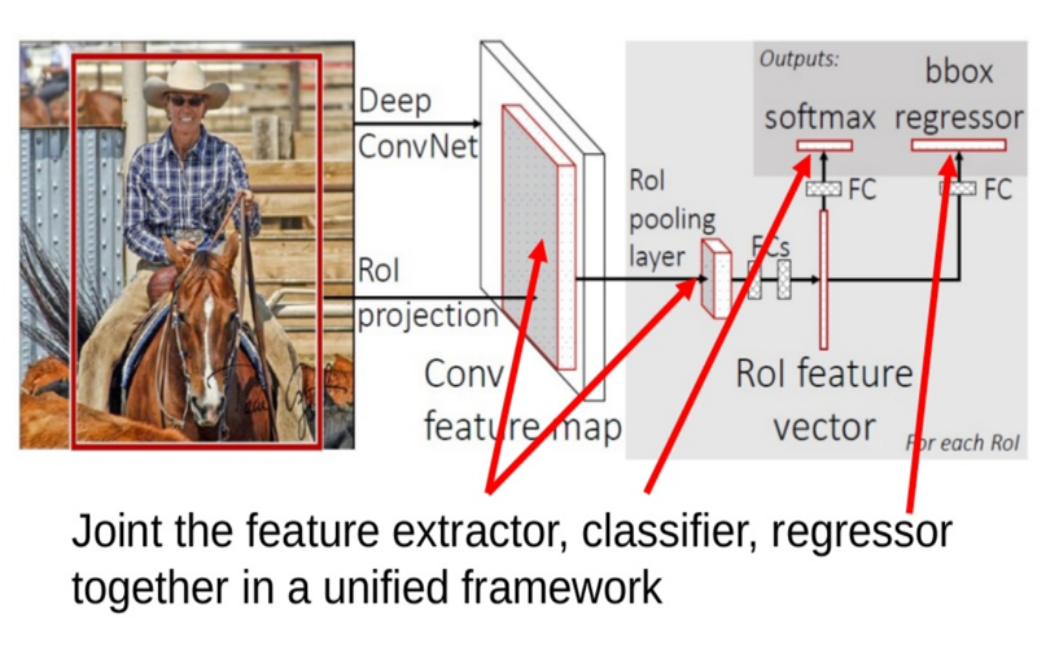


# Fast R-CNN 선정 이유

## Fast R-CNN의 해결책

1) ***RoI pooling*** 기법을 적용 -> 중복 연산 최소화 -> training speed 문제 극복

2) 세 개의 모델을 ***하나의 모델***로 통합 -> 연산 양 감소 -> performance 문제 극복



# 프로젝트 기대 효과

1) Object Detection 분야에서 상당히 좋은 평가를 받고 있는

**Fast R-CNN**을 malware classification 분야에 활용하였을 때도

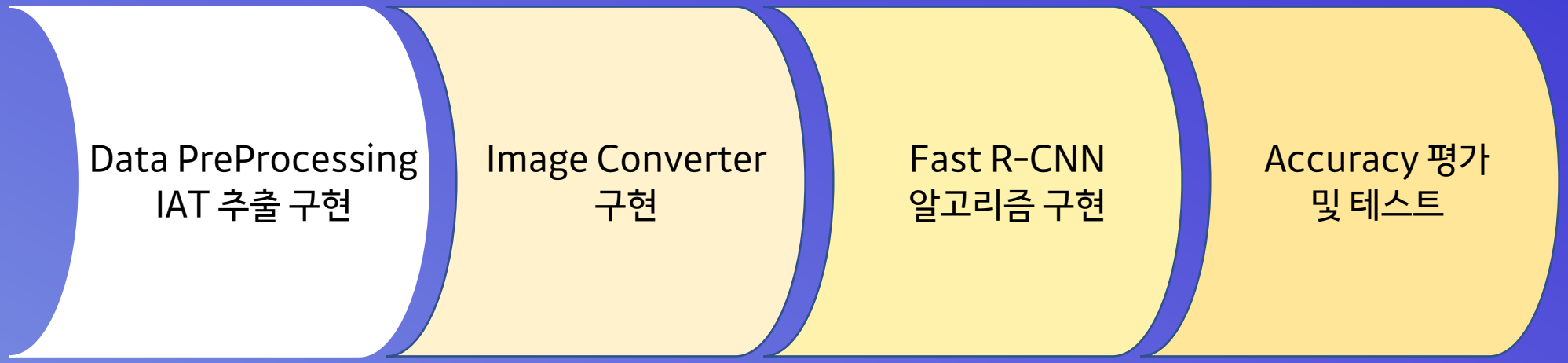
상당히 좋은 결과를 낼 것이라 기대

2) Binary Malware file 전체를 input으로 사용하는 것이 아닌,

IAT를 input으로 학습함으로써, **악성코드의 sequential 특징**을

학습할 수 있을 것이라고 기대

# 프로젝트 추진 계획


Data PreProcessing
IAT 추출 구현
Image Converter
구현
Fast R-CNN
알고리즘 구현
Accuracy 평가
및 테스트