딥러닝 모델의 취약점을 판단하기 위한 커버리지 기반의 퍼징 시스템

기술 개요
Overview

① 적용분야
인공지능 보안 취약점 분석 및 공격 탐지 시스템

② 기술요약
딥러닝 모델의 보안 취약점 분석을 통해 딥러닝 모델이 공격으로 인해 오동작 하는 것을 방지하기 위한 것으로, 원본 이미지를 변형하여 생성된 이미지에 대한 딥러닝 모델의 예측 결과와 원본 이미지의 클래스에 기초하여 딥러닝 모델의 예측 결과에 대한 오류를 자동 탐지함으로써, 사람의 개입 없이 딥러닝 모델의 취약점 검출이 가능하며 딥러닝 모델의 취약점 검출을 위해 소요되는 시간을 절약할 수 있음

③ 특허 권리 범위
- 원본 이미지에 대한 변형 이미지를 이용하여 딥러닝 모델의 뉴런 커버리지(neuron coverage)를 측정하고, 변형 이미지의 클래스에 대한 딥러닝 모델의 예측 결과와 원본 이미지의 클래스에 기초하여 딥러닝 모델의 예측 결과에 대한 오류를 검출하며, 딥러닝 모델의 속성을 기초로 취약점의 종류를 분류함
- 변형 이미지는 원본 이미지에 대한 의미 보존성을 유지하는 적대적 예시를 생성할 수 있는 변환 기법을 이용하여 생성됨

기술의 목적
딥러닝 모델의 취약점에 대한 공격에 대응하기 위해 퍼징(fuzzing) 기법을 딥러닝 모델에 적용하려는 연구가 이루어지고 있으나, 대부분 단일 오류 탐지 후 종료되는 단순한 구조이고 입력 값 변환 기법의 효율성, 딥러닝 모델 커버리지 측정의 정확성에 대한 증명이 미비하므로, 종래 기술에 비해 효율성과 정확성을 개선할 수 있는 방안이 요구됨

해결 방안
원본 이미지에 대한 변형 이미지를 이용하여 딥러닝 모델의 뉴런 커버리지(neuron coverage)를 측정하고, 변형 이미지의 클래스에 대한 딥러닝 모델의 예측 결과와 원본 이미지의 클래스에 기초하여 딥러닝 모델의 예측 결과에 대한 오류를 검출하여 딥러닝 모델의 속성을 기초로 취약점의 종류를 분류함

기술의 특장점
사람의 개입 없이 자동으로 딥러닝 모델의 취약점 분석을 위한 변형 이미지 생성, 취약점 분석 및 뉴런 커버리지 측정이 가능하도록 함으로써 뉴런 커버리지 측정과 취약점 검출의 효율성과 정확성을 개선할 수 있음

기술적용 시 기업의 이점

딥러닝 모델 취약점 분석을 위해 요구되는 비용과 시간을 절약하면서 정확한 분석이 가능하므로 시장 경쟁력 확보 가능함

SWOT분석
Analysis

S 강점

사람의 개입 없이 자동으로 딥러닝 모델의 취약점 분석이 가능하므로, 취약점 분석을 위한 효율성을 향상시킬 수 있으며, 뉴런 커버리지 측정과 취약점 검출의 효율성과 정확성을 확보할 수 있음

W 약점

딥러닝 모델의 유형과 딥러닝 모델의 취약점에 대한 공격 유형이 다양해지고 있으나, 모든 모델 유형과 공격 유형에 대한 대응이 어려움

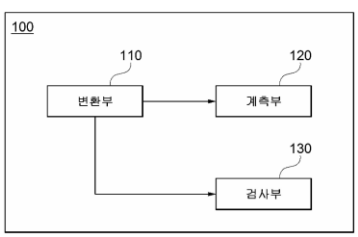
O 기회요인

딥러닝 기술의 적용 분야와 시장 수요가 증가하고 있으나, 딥러닝 모델이 가진 보안 취약점 위협을 방어하기 위한 기술에 대한 연구는 미흡한 상황이므로, 딥러닝 모델의 보안 취약점 분석 기술에 대한 시장 수요가 증가하고 있음

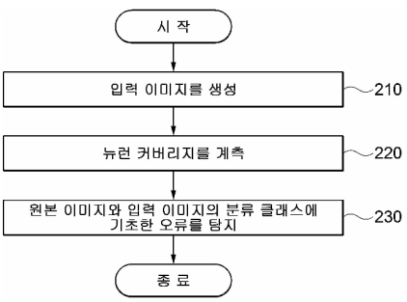
T 위험요인

국내 취약점 분석 시장 협소하며, 취약점 분석 업체들 사이의 가격 경쟁 심화되고 있음

대표도면
Drawing



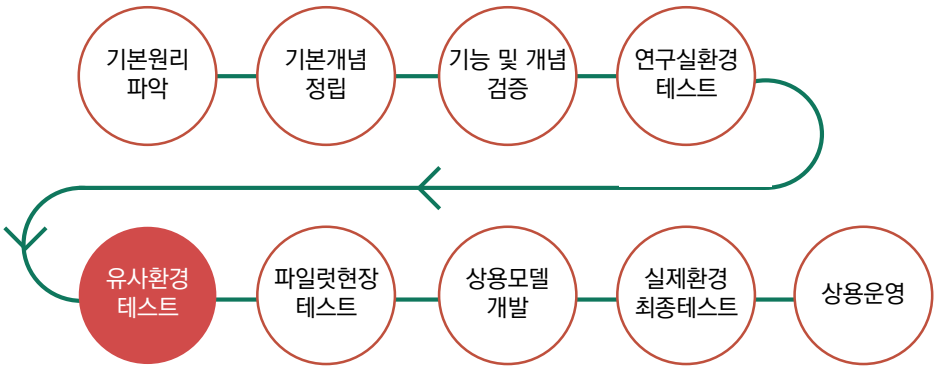
〈 취약점 판단 장치의 블록도 〉



〈 취약점 판단 방법에 대한 흐름도 〉

기술의 완성도
Technology Readiness level

● : 현재 단계입니다.



특허현황
Patent status

발명의 명칭	출원번호	등록번호	출원국가
딥러닝 모델의 취약점 판단 장치 및 방법	10-2020-0087811 (2020.07.15.)	10-2191722 (2020.12.10.)	한국

기술키워드
Keyword

한글키워드	영문키워드
딥러닝, 인공지능, 취약점, 퍼징	deep learning, AI, vulnerability, fuzzing

발명자
Inventor Info.

교수명	윤주범
소속	세종대학교 정보보호학과
연구분야	소프트웨어 취약점 분석, 네트워크 보안, AEG(Automatic Exploit Generation)
E-mail	jbyun@sejong.ac.kr
웹사이트	http://home.sejong.ac.kr/~jbyun/