

캡스톤디자인 II

응용레벨 암호화 트래픽 AI 기반 분석 기술 개발

2022390790

조윤지

Contents

01 | 연구 선정 배경

03 | 연구 과정

05 | 기대효과 및 보완할 점

02 | 캡스톤디자인 II 연구 수행 내용

04 | 연구 결과

연구 선정 배경



연구 배경 및 필요성 정리

암호화 트래픽 증가,
기존 분류 방식의 한계 확인



기존 연구 분석

공개 암호화 트래픽 데이터셋 확인,
초기 전처리 설계



데이터셋 탐색 및 전처리 기초

공개 암호화 트래픽 데이터셋 확인,
초기 전처리 설계



기본 모델 적용 실험

기초 정확도 확보

캡스톤디자인 II 연구 수행 내용

1. 데이터 확장 및 전처리 고도화



다양한 암호화 트래픽 데이터셋 확보 및 통합
노이즈 제거, 특징 추출 기법 적용으로 학습 효율 향상
학습용/검증용 데이터셋 분리 및 구조 정제

3. 실시간 분석 파이프라인 구축



트래픽 캡처 → 전처리 → 모델 추론 → 결과 시각화 과정 자동화
모듈별 인터페이스 설계 및 연동
실시간 모니터링 가능 대시보드 구현

2. 딥러닝 모델 최적화



CNN, RNN, Transformer 기반 구조 비교 및 적용
하이퍼파라미터 조정으로 정확도 및 재현율 향상
모델 경량화로 추론 속도 개선

4. 시스템 통합 및 검증



데이터 처리·모델·시각화 모듈을 통합 시스템으로 구성
다양한 네트워크 환경에서 성능 테스트 진행
처리 속도, 정확도, 안정성 검증 및 문제점 보완

연구 과정

CNN 기반 암호화 트래픽 분석 모델 구현을 목표로 데이터 전처리와 기본 모델 구조를 세팅

데이터셋 확보

Flow-level Synthetic Dataset 생성 및 구조 정의

전처리

Flow-level 특징 정규화 및 분할
StandardScaler를 활용한 특징값 정규화
Train / Validation / Test 데이터셋 분할
CNN 입력 형식(샘플 수 × 특징 수 × 1)으로 reshape

모델구성

1D CNN 기반 분류 모델 설계

학습 결과

학습 성능 확인 및 튜닝 진행
Epoch 12 기준 Train/Validation Accuracy 99% 이상 수렴
Loss 0.01 이하로 안정적으로 감소

시각화

CNN 모델 성능 및 평가 시각화

연구 결과

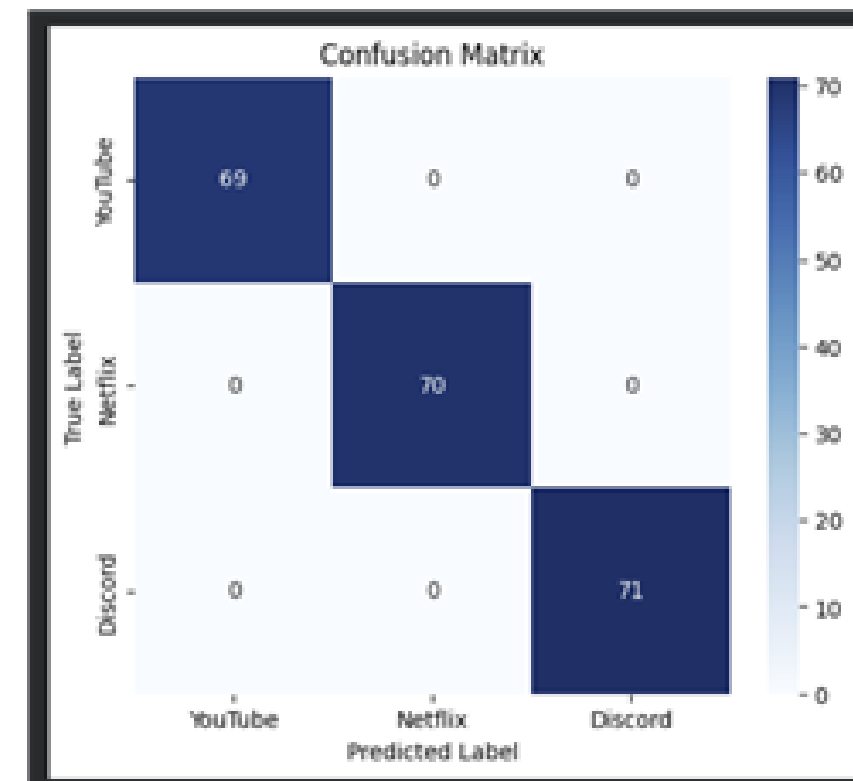
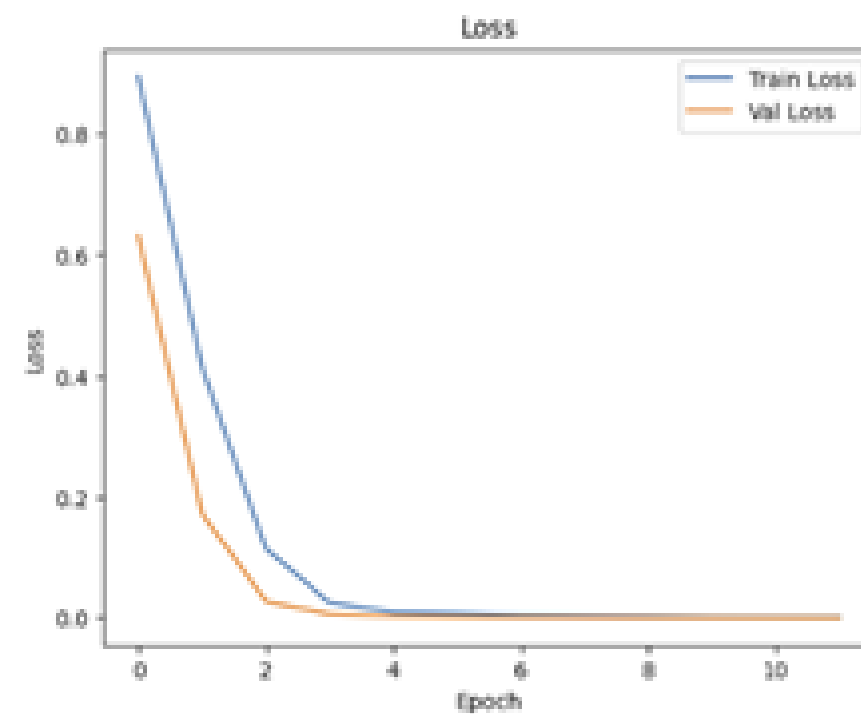
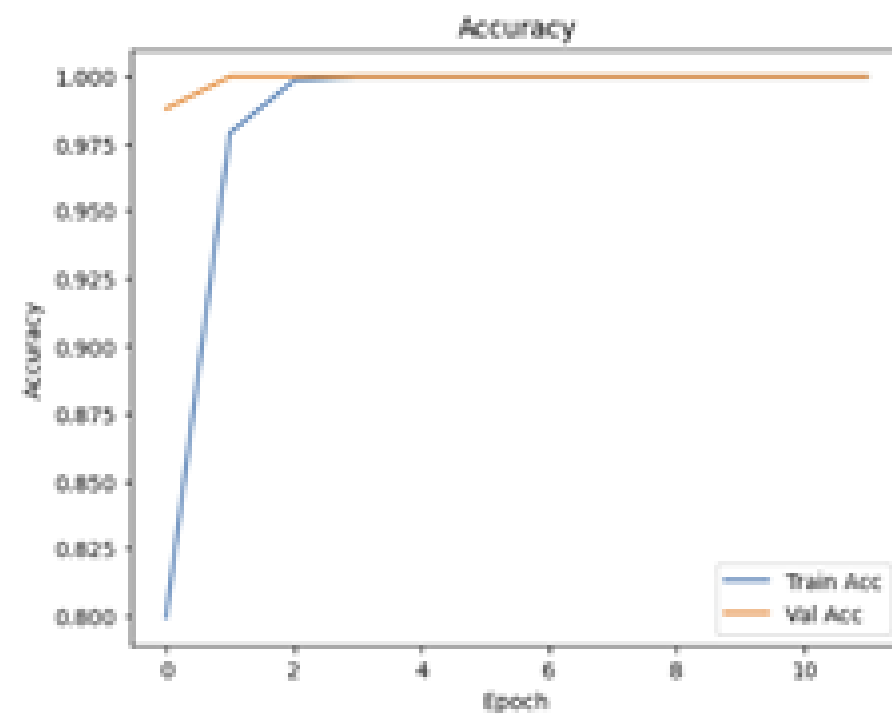
1) 딥러닝 학습 성능

Accuracy: 약 99.5% 이상

F1-score: 약 99% 이상

Train/Validation Loss 모두 3 epoch 내 안정적 감소

Confusion Matrix: YouTube/Netflix/Discord 오분류 없음



연구 결과

2) 최종 시스템 구성 결과

CNN 모델: Flow-level 기반 고정밀 분류

C++ 추론기: CSV 입력 기반 실시간 분석 가능

전체 파이프라인:

“데이터셋 → CNN 학습 → 평가 → C++ 추론기 변환 → 데모 실행”

3) 데모 실행

임의 트래픽 입력 → 예측 결과 출력

YouTube/Netflix/Discord 정확하게 분류됨

실제 환경 적용 가능성을 확인

기대 효과

01 | 암호화 트래픽 분석 분야에서 사용되는 딥러닝 모델 설계-평가-시스템 구현까지의 실무형 경험 확보

02 | Flow-level 기반 분석이 실제 기업 보안 모듈에서도 사용 가능성을 확인

보완할점

- 01 실제 PCAP 기반 실데이터 수집 및 추가 실험 필요
- 02 noise 포함 환경에서 robustness 평가 필요
- 03 클래스 다양화(Zoom, KakaoTalk 등) 확장할 필요
- 04 실시간 패킷 캡처 기능 및 간단한 UI 추가 개선 가능