PORTFOLIO

이름	주 상 현
생년월일	1994. 09. 27
이메일	kidney94@naver.com
전화번호	010-7172-3817

Profile

이름 : 주상현

- **한성대학교 지능형 컴퓨팅랩**

✓ 지도교수 : 김명선

- **연구 분야**

✓ 인공지능 컴퓨팅 및 응용

✓ Vision AI
(Adversarial Attack Defense & Detect)

✓ 리눅스 시스템 최적화

✓ 우선순위 GPU 스케줄링

✓ 임베디드 시스템 SW

게재 논문

- **SCIE 저널**

✓ Time-Constrained Adversarial Defense in IoT Edge Devices through Kernel Tensor Decomposition and Multi-DNN Scheduling

➢ **MDPI Sensors, 2022.08.**

- **KCI 저널**

✓ 소프트웨어 기반 임베디드 시스템용 적대적 공격 검출 시스템

➢ **대한전자공학회 논문지, 2022.07.**

PROJECT

- **2022.03. ~ 2022.11.**

✓ 중대형 공간용 초해상도 비정형 플렌옵틱 동영상 저작/재생 플랫폼 기술 개발

SKILL

- **C/C++, Python, Torch, CUDA**

PORTFOLIO

CONTENTS

Paper 01.

Time-Constrained Adversarial Defense in IoT Edge Devices through Kernel Tensor Decomposition and Multi-DNN Scheduling

About paper

✓ 본 논문은 **비 선제적 장치인 임베디드 GPU**에서 여러 인공지능 어플리케이션이 동시에 실행되고 있을 때 **적대적 예제 복원 작업이 빠르고 우선적으로 실행**될 수 있도록 하는 **GPU 스케줄링 프레임워크**를 제안함.

- **국외 SCIE (MDPI Sensors)**
- **게재일 : 2022.08.07**

Paper 01. Time-Constrained Adversarial Defense in IoT Edge Devices through Kernel Tensor Decomposition and Multi-DNN Scheduling

Motivation

- 적대적 예제 : DNN의 오분류를 유도하는 noised Image
- 적대적 예제 복원 과정 : 적대적 예제 →적대적 예제 복원 → DUNET(복원 모델) → 복원된 이미지 → Inception-V3(target DNN)
- 임베디드 디바이스에서 다수의 DNN이 단일 GPU에서 동시에 실행될 때 복원 과정을 빠르고 우선으로 실행하고자 함

Problems

- 임베디드 디바이스에서 DUNET의 단일 데이터 추론 시간이 다른 DNN에 비해 오래 걸리는 것을 확인

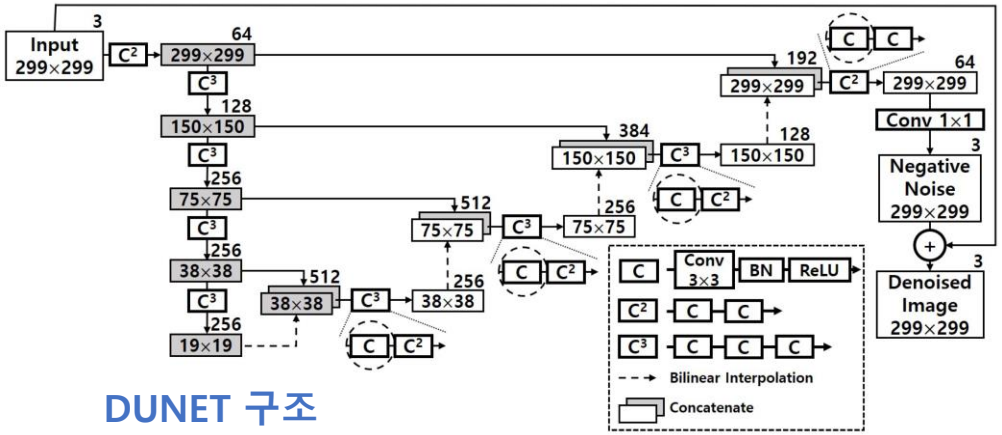
	Inception-V3	ResNet-152	VGG-16	DUNET
Jetson AGX Xavier	86.3 ms	157.7 ms	50.4 ms	160.7 ms
Jetson TX2	103.9 ms	198.6 ms	76.8 ms	246.4 ms

- GPU는 우선순위에 따른 실행이 불가능한 프로세서

Paper 01. Time-Constrained Adversarial Defense in IoT Edge Devices through Kernel Tensor Decomposition and Multi-DNN Scheduling

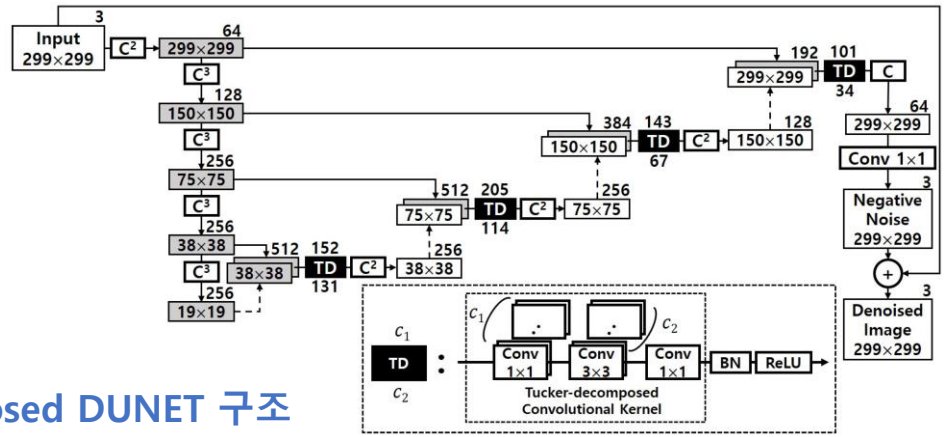
▪ Solution

➢ DUNET의 연산량을 줄이기 위해 DUNET에 Convolution Kernel Decomposition 경량 기법 적용



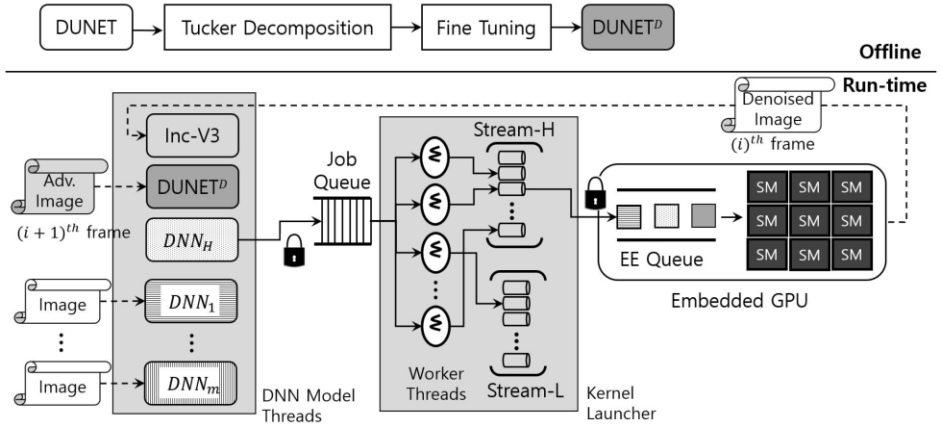
DUNET 구조

경량



Decomposed DUNET 구조

➢ 우선순위에 따른 실행이 가능한 GPU scheduler 제안



- **Offline** : DUNET 경량 후, fine tuning하여 DUNET^D 생성
- **Run-time** : 크게 DNN Model Threads, Job Queue, Kernel Launcher로 구성

- **DNN Model Threads** : 각 DNN을 실행하는 thread 생성, 각 DNN은 vector로 재구성하여 Job의 단위를 다르게 설정
- **Job** : GPU에 DNN을 할당하는 단위 ex) [conv, batch_norm, relu] or [conv]
- **Job Queue** : 우선순위 Queue로써, 순위가 높은 Job을 먼저 배출
- **Kernel Launcher** : 우선순위가 높은 job은 Stream-H로 전달되어 GPU의 EE Queue에 먼저 전달됨

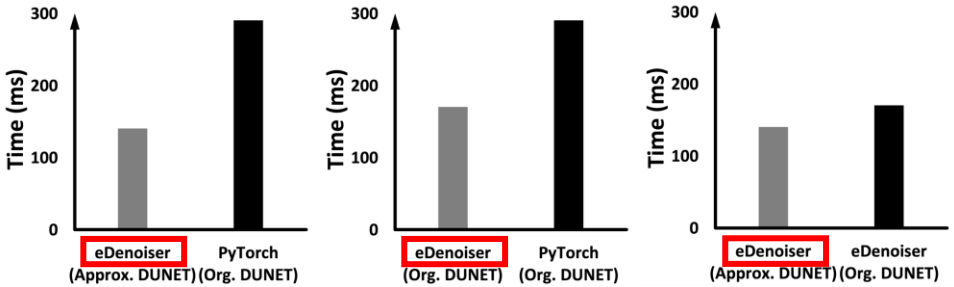
Time-Constrained Adversarial Defense in IoT Edge Devices through Kernel Tensor Decomposition and Multi-DNN Scheduling

Result

➢ DUNET^{D}의 정확도 감소율이 미미함 (최대 1.78%)

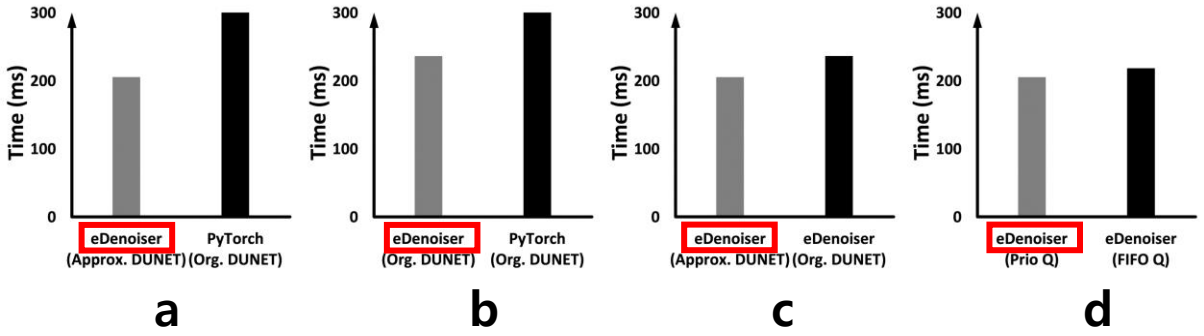
	Result in [8]	Org. DUNET	Approx. DUNET
Clean-image-test-set	76.2%	76.53%	76.38%
White-box-test-set	75.2%	72.37%	70.59%
Black-box-test-set	75.1%	74.86%	74.45%

➢ DUNETD와 Inception-V3(target DNN)만 실행하여 경량화로 인한 속도 향상과 GPU scheduler 효능 확인



- **a** : 경량화와 GPU scheduler 적용 전후 비교 (51.72% 속도 향상)
- **b** : GPU scheduler 적용 전후 비교 (41.3% 속도 향상)
- **c** : 경량화 적용 전후 비교 (17% 속도 향상)

➢ ResNet-152, RegNet, ResNext, WideResNet과 동시에 복원 과정을 실행했을 때 효능 확인



- **a** : 경량화와 GPU scheduler 적용 전후 비교 (48.36% 속도 향상)
- **b** : GPU scheduler 적용 전후 비교 (40.55% 속도 향상)
- **c** : 경량화 적용 전후 비교 (13.13% 속도 향상)
- **d** : 우선순위 queue 적용 전후 비교 (6% 속도 향상)

Paper 02.

소프트웨어 기반 임베디드 시스템용 적대적 공격 검출 시스템

About paper

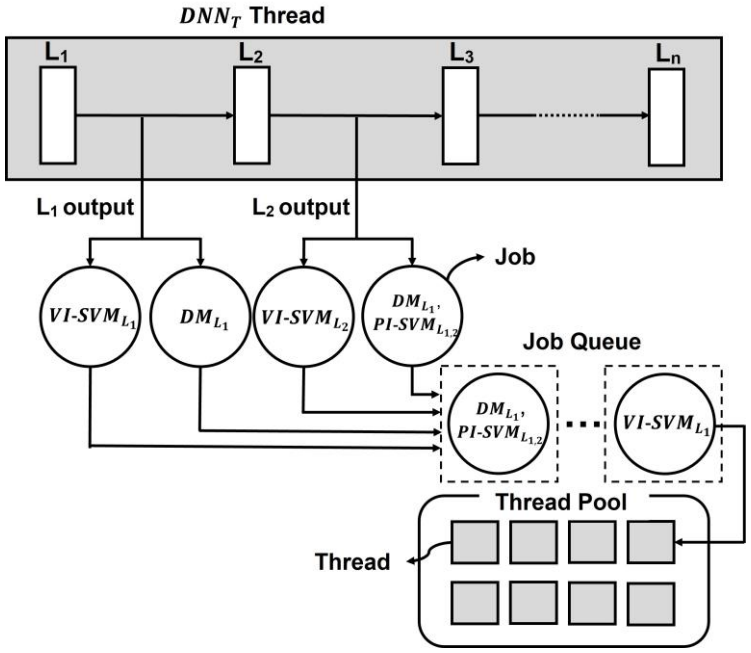
✓ 본 논문은 성능은 좋지만 메모리 사용량이 크고 수행 시간이 오래 걸리는 적대적 공격 탐지 기법인 **NIC**(Neural Network Invariant Checking)를 임베디드 시스템에서 target DNN과 동시에 실행할 수 있도록 개량한 **소프트웨어 기반 시스템**을 제안함.

- **국내 SCIE (대한전자공학회)**
- **게재일 : 2022.07.25**

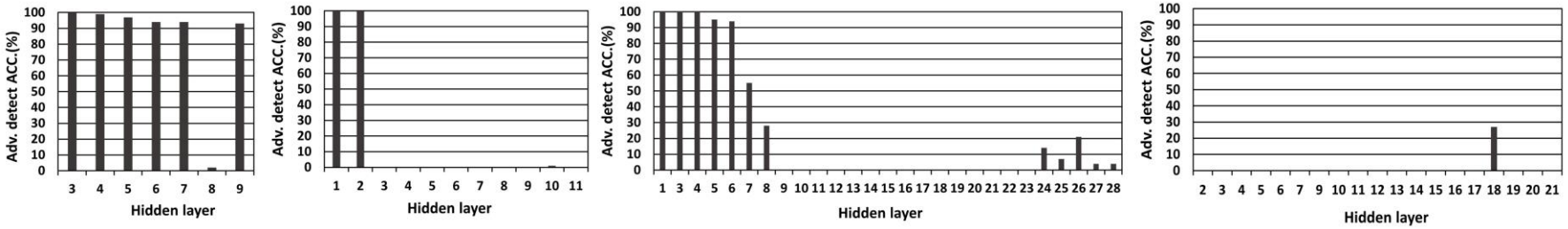
소프트웨어 기반 임베디드 시스템용 적대적 공격 검출 시스템

- **GOAL**
  - 본 논문은 임베디드 시스템(Jetson AGX Xavier)에서 하드웨어 변경 없이 NIC와 target DNN을 동시에 실행하여 두 개의 수행 종료 시간 차이를 최소화하고 메모리 사용량을 최소화 하고자 함

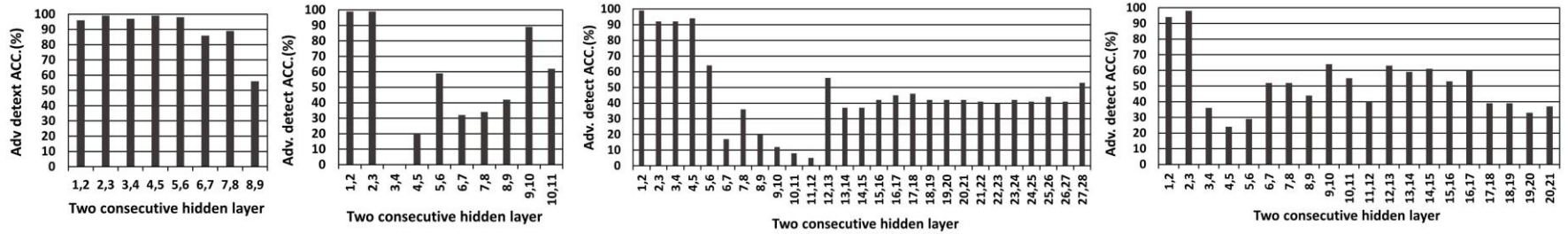
- **Contributions**
  - Multi Thread 기반 시스템을 설계하여 GPU와 CPU의 동시 사용량을 최대화하여 NIC와 target DNN의 수행 시간 차이를 좁힘
  - target DNN의 특정 layer에서만 공격 탐지를 진행하여 target DNN별 메모리 사용량을 최소화함



Workflow of proposed system

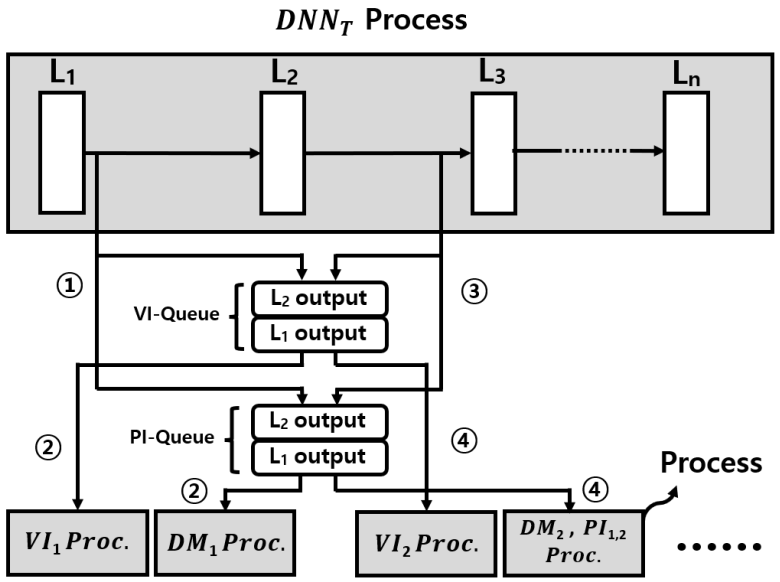


VI violation detection rate of adversarial examples in each hidden layer of the target DNN



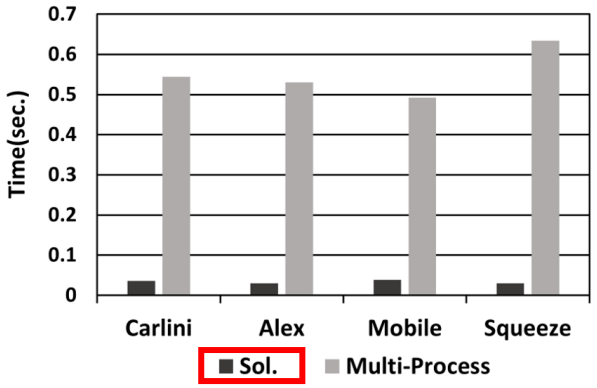
PI violation detection rate of adversarial examples in each hidden layer of the target DNN

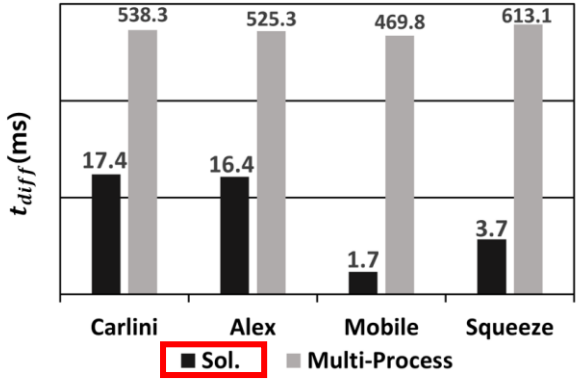
- **Result**
  - ➢ 제안한 시스템의 성능을 확인하기 위해 기존 멀티 프로세스 환경을 직접 구현

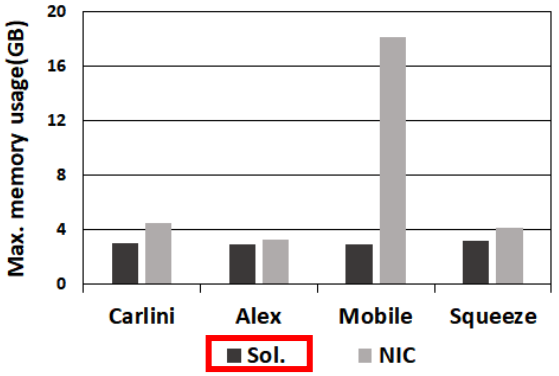


✓ t_{diff} : target DNN 추론 완료 시간 – 공격 탐지 종료 시간

  - ➢ 제안한 시스템 적용 전 대비 실행 시간 최대 95.2%, t_{diff} 최대 99.6%, 메모리 사용량 최대 83.9% 감소







Project 01.

중대형 공간용 초고해상도 비정형 플렌옵틱 동영상 저작/재생 플랫폼 기술 개발

About project

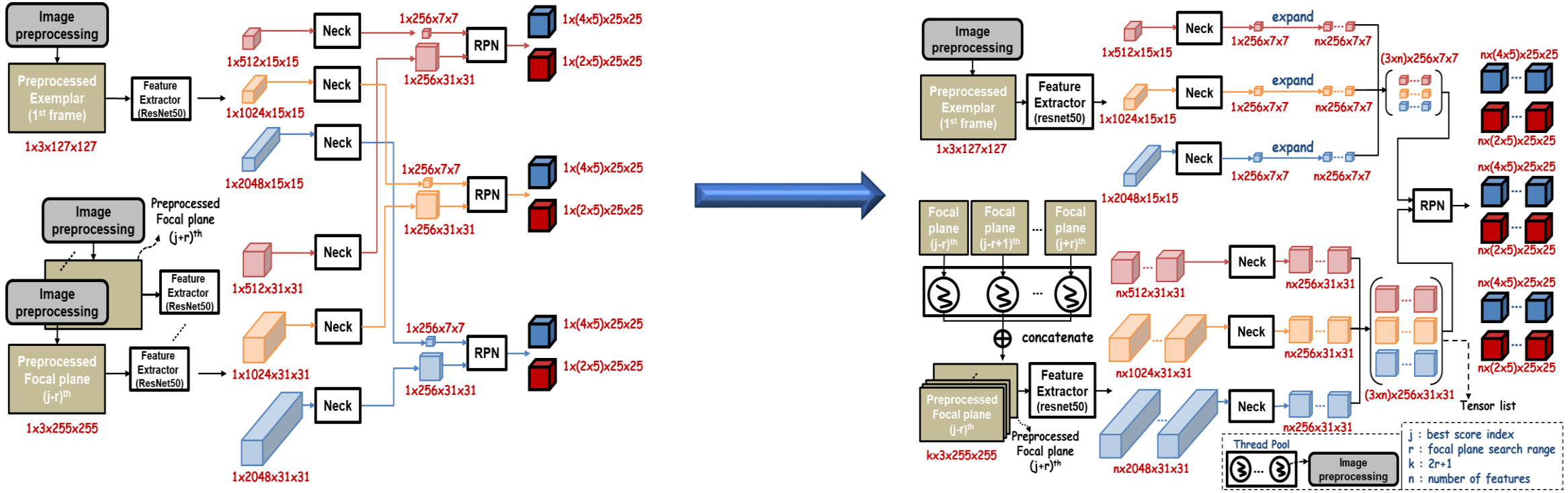
✓ 본 연구 개발의 최종 목표는 CPU 및 GPU 구조에 기반하여 병렬화 및 최적화된 **객체 추적 알고리즘**을 구현함으로써 **플렌옵틱 영상 시퀀스 입력에 대한 고속/고정밀의 단일 및 다중 객체 추적이 가능한 기술을 개발**하기 위함

- **with Hansung University & ETRI**
- **Mar. 2022 ~ Nov. 2022**

Project 01. 중대형 공간용 초고해상도 비정형 플렌옵틱 동영상 저작/재생 플랫폼 기술 개발 (Mar. 2022 ~ Nov. 2022)

- **GOAL**
  - Python으로 구현된 Siamrpn++의 플렌옵틱 영상 다중 객체 추적의 속도 향상 및 최적화 (4sec/frame)

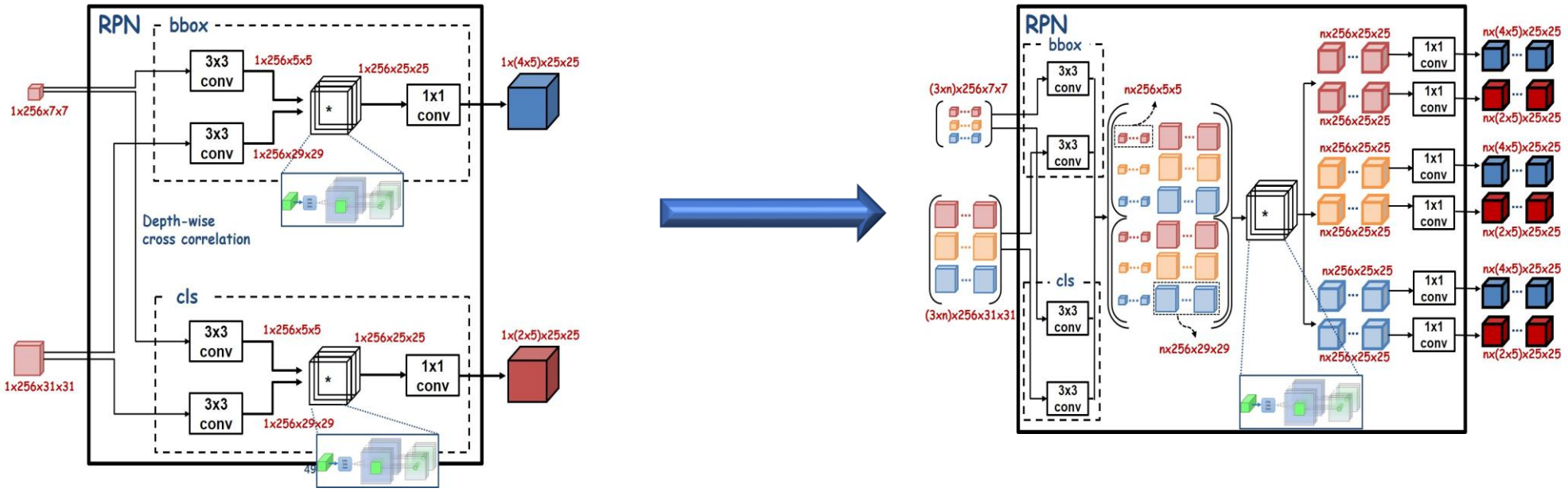
- **Contributions**
  - 이미지 전처리 과정 멀티 스레드화
    - ✓ 하나의 프로세스 내에서 여러 개의 이미지 전처리 스레드를 생성하여 CPU에서 동시에 여러 포컬플레인에 대한 이미지 전처리 진행



Project 01. 중대형 공간용 초고해상도 비정형 플렌옵틱 동영상 저작/재생 플랫폼 기술 개발 (Mar. 2022 ~ Nov. 2022)

Contributions

- 포컬플레인을 batch 단위로 입력 받음
  - ✓ Resnet-50을 통한 feature 추출이 batch 크기에 비례하여 횟수가 감소됨
  - ✓ RPN 당 Depth-wise cross correlation 연산 회수를 2번에서 1번으로 줄임



Result

- 추적 속도가 19.6% 감소하여 4sec/frame을 충족하는 3.41sec/frame 달성

실행 version	sec/frame
Python version	4.24
Solution version	3.41