Data-driven AI Security HCI Lab (DASH-Lab):

Developing Usable and Secure Technology after Better Understanding Data, Machines, and Humans

Professor : Simon S. Woo

우사이먼성일 교수

College of Computing and Informatics, Sungkyunkwan University

Department of Artificial Intelligence, Sungkyunkwan University

Department of Applied Data Science, Sungkyunkwan University

Researcher Recruit

학석연계/석사/박사, 석•박 통합 및 **DASH-Lab**에서 연구자로서 성장하기를 희망하는 학생을 모집합니다. 저희 연구실에서는 보안 및 개인정보, **HCI, AI/Machine Learning**를 주로 연구하고 있으며 관심이 있으신 분들은 우사이먼성일 (**swoo@g.skku.edu**) 교수님께 연락주세요. 더 많은 정보는 https://dash-lab.github.io/ 에서 확인하실 수 있습니다.

Abstract

DASH-Lab은 여러 **Machine Learning (ML)** 및 **Artificial Intelligence (AI)**를 기반으로 하여 **Object Detection, Deepfake Detection, Time Series, and Model Compression** 등의 실생활에서 다양하게 발생하는 문제들에 도전합니다.

현재 연구실에서 다루고 있는 주요 주제들은 아래와 같습니다.

1) **Image translation**을 통한 데이터 증강을 활용하여 다양한 시나리오에서의 쓰러진 사람 탐지.
2) **Oriented bounding box**를 활용한 위성 이미지의 작은 객체 탐지.
3) 멀티 생성 도메인에서의 딥페이크 이미지 탐지.
4) 시계열 이상치 탐지.
5) 프루닝(**Pruning**) 및 지식 증류 (**Knowledge Distillation**)을 활용한 가볍고 효율적인 모델 개발.
6) 실생활 문제를 해결하기 위한 **robust** 벤치마크 데이터셋 생성.
7) 개인정보보호 관련 정책 변화를 유연하게 반영하여 준수하는 **AI** 플랫폼 개발
8) **GAN**을 활용한 의료 이미지 생성.
9) **AI**를 활용한 **Safety Signals of Drugs** 탐지

이외에도, 저희 연구실에서는 학부연구생 및 인턴들에게 국내 및 해외 학회 게재를 목표로 개개인의 연구나 연구를 위한 프로젝트 참여를 권장하고 있습니다. 더 많은 정보는 다음 기사에서 확인하실 수 있습니다 .

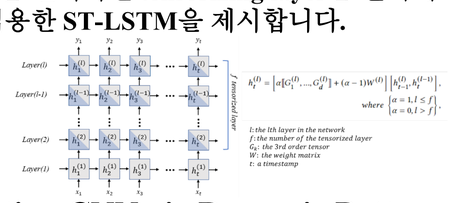
https://www.skku.edu/skkuzine/section/culture03.do?articleNo=96752&pager.offset=0&pagerLimit=10

NEWS

DASH 연구실에서는 인공지능 및 정보검색 분야의 **top-tier** 국제학술대회인 **CIKM(Conference on Information and Knowledge Management) 2022**에 **full paper** 논문 **5**편, **workshop** 논문 **1**편이 게재 승인되어 **10**월에 발표될 예정입니다.

◆ Selective Tensorized multi-layer LSTM (ST-LSTM)

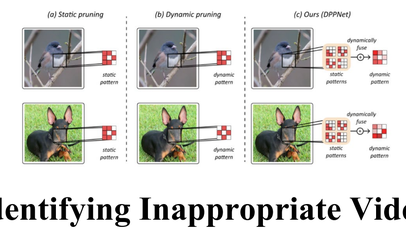
➢ 본 연구에서는 위성의 궤도를 예측하는 모델로서, 딥러닝의 웨이트 매트릭스를 텐서화한 **tensorizing layer**를 멀티레이어 **LSTM**에 선택적으로 적용한 **ST-LSTM**을 제시합니다.



◆ Accelerating CNN via Dynamic Pattern-based Pruning Network

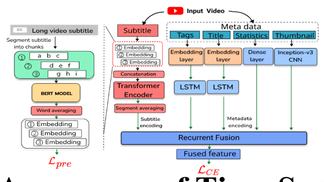
➢ 기존의 **dynamic pruning method**는 가속을 위해 수행되는 추가적인 오버헤드때문에 실제 가속이 이루어지기 어려웠습니다.

➢ 본 논문에서는 실제 가속이 가능한 **dynamic pruning method**를 제안합니다.



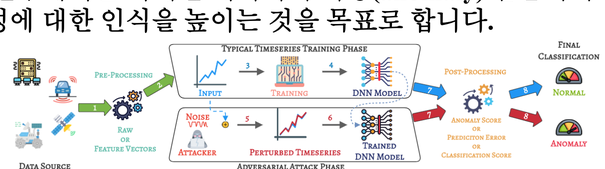
◆ Samba: Identifying Inappropriate Videos for Young Children on YouTube

➢ 본 논문에서는 어린이용 유튜브 동영상을 분류하기위해 메타데이터와 비디오 자막을 모두 사용하는 퓨전 모델을 제안합니다.



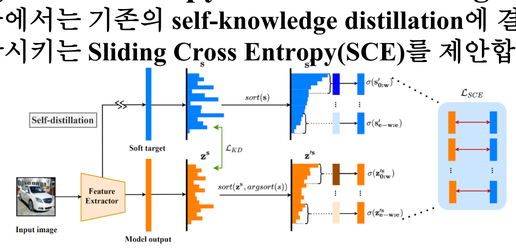
◆ Towards an Awareness of Time Series Anomaly Detection Models' Adversarial Vulnerability

➢ 본 연구에서는 시계열 데이터의 이상(**anomaly**) 검출기의 적대적 취약성에 대한 인식을 높이는 것을 목표로 합니다.



◆ Sliding Cross Entropy for Self-Knowledge Distillation

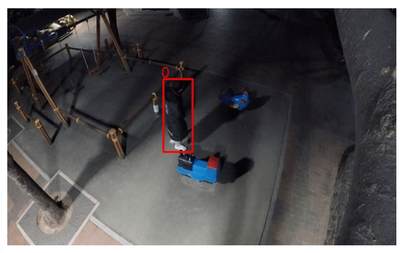
➢ 본 연구에서는 기존의 **self-knowledge distillation**에 결합하여 성능을 향상시키는 **Sliding Cross Entropy(SCE)**를 제안합니다.

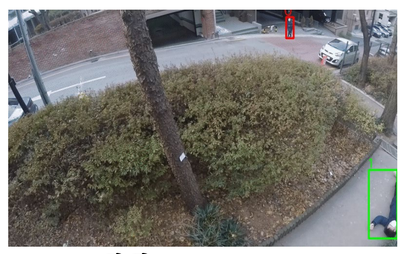


Object Detection

◆ AI Grand Challenge (AGC)

DASH-Lab은 **2020**년부터 **MSIT**에서 주관한 **AI Grand Challenge (AGC)**에 참여해 왔습니다. 이 대회의 주 목적은 실시간으로 주어진 비디오에 등장하는 쓰러진 사람을 탐지하는 것입니다. 탐지 시스템과 보간 알고리즘을 활용하여 **DASH-Lab**은 여러 기업과의 경쟁에서 **1**등을 달성하였고 장관상을 수여 받았습니다. 추후에 있을 (2022년 말) 마지막 대회에서 더욱 다양한 시나리오를 기반으로 엣지 디바이스에서의 추론 성능을 보이는 과제를 목표로 합니다.





◆ Aerial Image Object Detection (proj. with 한화)

위성 이미지 탐지는 매우 큰 위성 이미지로부터 위치와 객체의 카테고리를 탐지 및 추정하는 것을 목표로 합니다.

풀어야 할 문제는 크게 아래 **2**가지로 요약할 수 있습니다.

➢ 위성 이미지 객체의 극도로 작고 조밀한 분포.

Horizontal bounding box를 기반으로 하는 일반적인 객체 탐지와는 달리 위성 이미지는 우주에서 촬영되었기 때문에 객체들이 이미지의 축과 평행할 가능성이 낮습니다. 이 과정에서 객체 중심의 **oriented bounding boxes**를 활용합니다.

➢ 초고해상도 입력 이미지.

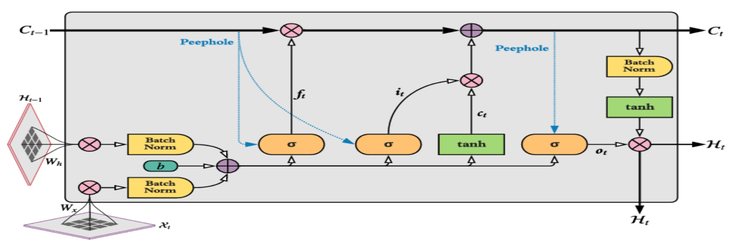
위성 이미지의 크기는 매우 크기 때문에, 이 과제에서 객체를 탐지하는 데에 있어 이미지의 패치화는 필수입니다. 따라서 실제로는 아주 큰 객체라도 굉장히 먼 거리에서 촬영되었기 때문에 패치화 단계에서 잘리는 현상이 발생하고, 이를 보강하는 알고리즘을 개발하는 것이 핵심입니다.

Deepfake Detection

◆ CLRNet

➢ 현존하는 딥페이크 탐지 알고리즘은 훈련된 데이터셋이 아닌 다른 딥페이크 이미지를 탐지 하는 데에 있어 어려움이 있습니다.

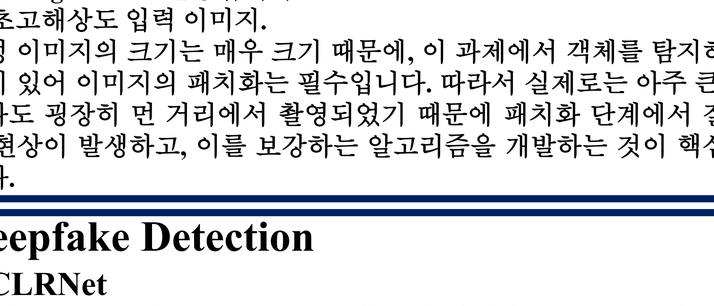
➢ 본 연구에서는 독창적인 학습 전략과 공간적 정보 뿐만 아니라 시간적 정보를 동시에 학습할 수 있는 **Convolutional LSTM-based Residual Network, CLRNet**을 제시합니다.



◆ ADD

➢ 낮은 해상도의 이미지는 **1) fine-grant artifact**를 제거하는 loss of high-frequency components, 2) loss of correlated information의 두가지의 고질적인 문제를 갖고 있습니다.

➢ 이러한 문제를 완화하기 위하여 본 연구에서는 **Knowledge Distillation**을 적용하여 두개의 새로운 증류 모듈: **frequency attention distiller & multi-view attention distiller**을 제시합니다.



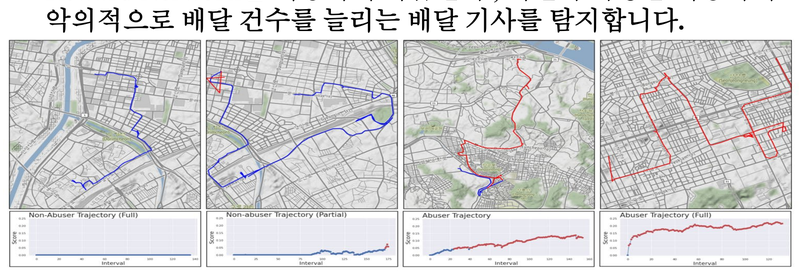
◆ CoRED

➢ **CoRED**는 **Continual Learning, Knowledge Distillation, Representation**을 활용하여 여러 딥페이크 알고리즘으로부터 생성된 가짜 이미지를 효율적으로 학습할 수 있습니다.

Time Series

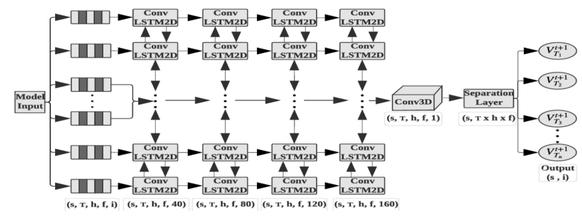
◆ Detecting Food Delivery Abusers using Variational Reward Inference Networks

➢ 정상적인 차량을 이용하는 배달 기사들의 **trajectory**를 **Variational Reward Nerwork**를 이용하여 학습한 후, 비인가 차량을 이용하여 악의적으로 배달 건수를 늘리는 배달 기사를 탐지합니다.



◆ Multivariate Convolution LSTM with Mixtures of Probabilistic PCA

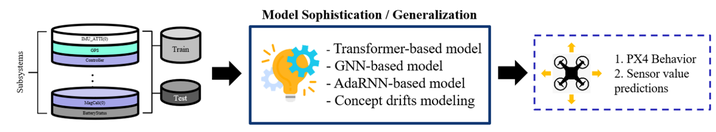
➢ 이상징후 탐지 성능을 높이기 위해, 이 연구에서는 뉴럴 네트워크와 **probabilistic clustering**을 함께 사용하여 **Multivariate Convolution LSTM with Mixtures of Probabilistic Principal Component Analyzer**를 구축합니다.



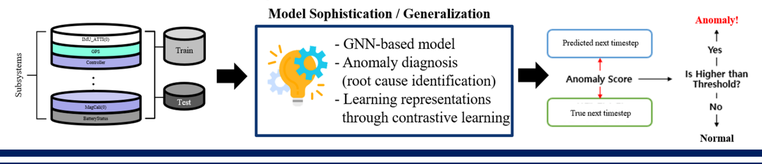
◆ Drone Multivariate Time Series Data

➢ 드론의 대규모 상용화에 따른 안전성·보안성의 중요성이 커짐에 따라, 우리는 드론 데이터에서의 시계열 데이터 예측 및 이상 징후 탐지에 대해 연구하고 있습니다.

➢ **Time Series Forecasting**



➢ **Time Series Anomaly Detection**

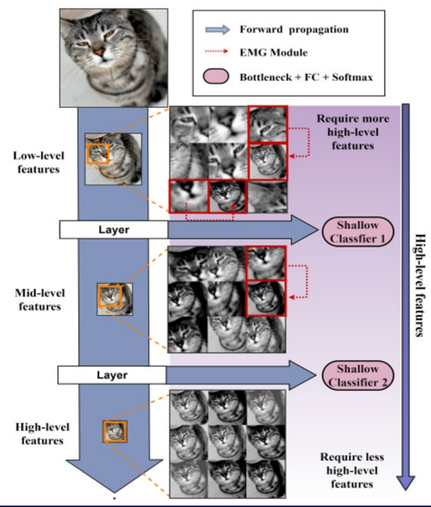


Model Compression

◆ EMGNet

➢ 기존의 **Early Exit Network**들은 **computational cost (FLOPs)**를 줄이기 위해 고안되었으나, **shallow classifier**내의 **high-level feature**의 부족으로 인하여 성능 감소 폭이 매우 크다.

➢ 해당 연구에서, 우리는 **auxiliary classifier**의 성능을 향상시키기 위해, **early exit network**에서 **multi-scale feature**를 생성하는 새로운 프레임워크인 **EMGNet**을 제안한다.



관련 프로젝트 협력 기관

SAMSUNG SDS

KISA Korea Internet & Security Agency

NSR 국가보안기술연구소 National Security Research Institute

NRF 한국연구재단

KARI 한국항공우주연구원

Hanwha

Acknowledgment

Our research is supported by IITP, NRF, SKKU, ICT, the Ministry of Science, ICT, Hanhwa System, and Samsung SDS.

IITP