

실시간 비정형객체 인식 기법 기반 지능형 이상 탐지 시스템에 관한 연구

이석창[†], 김영현^{**}, 강수경^{***}, 박명혜^{****}

Research on Intelligent Anomaly Detection System Based on Real-Time Unstructured Object Recognition Technique

Seok Chang Lee[†], Young Hyun Kim^{**}, Soo Kyung Kang^{***}, Myung Hye Park^{****}

ABSTRACT

Recently, the demand to interpret image data with artificial intelligence in various fields is rapidly increasing. Object recognition and detection techniques using deep learning are mainly used, and video integration analysis to determine unstructured object recognition is a particularly important problem. In the case of natural disasters or social disasters, there is a limit to the object recognition structure alone because it has an unstructured shape. In this paper, we propose intelligent video integration analysis system that can recognize unstructured objects based on video turning point and object detection. We also introduce a method to apply and evaluate object recognition using virtual augmented images from 2D to 3D through GAN.

Key words: Object Detection, Anomaly Detection, Class Activation Map, Unstructured Object Recognition, Gan, Transfer Learning, Self-Learning

1. 서 론

오늘날 딥러닝 학습 알고리즘 개선과 하드웨어 인프라 환경 개선 등에 힘입어 인공지능 영상 인식능력과 인지능력이 비약적으로 발전해오고 있다. 특히 의료, 제조, 자율주행, CCTV 등과 융합되어 빠른 성장세를 보이고 있고 넷플릭스, 유튜브, 인터넷 TV 등의 동영상 콘텐츠가 전체 트래픽에서 차지하는 비율이

2016년 73%에서 2021년 82%로 증가하는 등 영상 빅데이터의 증가세가 뚜렷하다[1]. 빅데이터 측면에서 보면, 현재 인류가 생산하는 데이터의 80%가 비정형 데이터이며, 비정형 데이터의 70%가 시각 빅데이터(이미지·동영상)이나, 내용 분석과 이해 기술은 텍스트(언어) 정보 대비 낙후된 분야로 중장기 관점에서 기술 확보가 필요하다. 스마트폰, CCTV, 블랙박스, 드론, 인공위성, 디지털 카메라 등에서 수집되는

* Corresponding Author : Myung Hye Park, Address: (34056) 105, Munji-ro, Yuseong-gu, Daejeon, Republic of Korea, TEL : +82-42-865-5251, FAX : +82-42-865-5269, E-mail : myunghye.park@kepco.co.kr

Receipt date : Jan. 17, 2022, Revision date : Feb. 9, 2022
Approval date : Feb. 11, 2022

[†] Power Wireless Communication Project Team, Electric Power Research Institute, Korea Electric Power Corporation (E-mail : leelang@kepco.co.kr)

^{**} Power Wireless Communication Project Team, Electric Power Research Institute, Korea Electric Power Corporation (E-mail : younghyun.kim@kepco.co.kr)

^{***} Power Wireless Communication Project Team, Electric Power Research Institute, Korea Electric Power Corporation
(E-mail : sk.kang@kepco.co.kr)

^{****} Power Wireless Communication Project Team, Electric Power Research Institute, Korea Electric Power Corporation

* This research was supported by the Industrial Drone Base Construction and Demonstration Project through the Korea Institute for Advancement of Technology (KIAT) supported by the Ministry of Trade, Industry and Energy (P0010318).

영상 데이터의 양은 기하급수적으로 증가하고 있으며, 이에 따른 비정형 영상 데이터를 인식하고 내용을 분석하여 활용할 수 있는 기술 요구가 점차 증가되고 있다. 사람의 인지체계는 80%가 시각에 의존하고 있으며 이미지·동영상 내용이해를 기반으로 하는 시각지식과 시각지능은 미래 성장을 견인할 동력원이 될 수 있으므로 시각 빅데이터 시대의 주도권 확보를 위해 “보여주는 영상”에서 “상황을 인지하는 영상”으로의 전환을 위한 기술이 필요하다. 그러기 위해서는 ‘영상수집 → 영상개선 → 영상분석, 상황인식 → 이벤트 확인 및 서비스 자동연계’의 과정으로 영상분석 솔루션 및 관제시스템과의 결합으로 보다 지능화된 플랫폼을 구축하여야 하며 IoT 기반의 네트워킹 환경, 엣지 분석, 클라우드 서비스를 통한 저장 환경, 웹/앱 서비스 솔루션 등의 추가적인 인프라 계층의 통합적인 플랫폼 구축이 필요하다. 영상수집과 영상분석 측면에서는 범죄·재난·재해 등 개인 및 사회의 안전위협 해소를 위해 CCTV의 확대설치 등의 국가적 관심도가 높은 반면 영상정보를 처리하는 방식이 여전히 노동집약적 수작업에 의해서 이루어지고 있으며 서로 다른 시스템에 의해 개별적으로 획득된 영상을 통합 분석하는 방법이 특히 부재한 편이다[2].

영상분석기술중 객체 인식에 있어 대표적인 딥러닝 객체 인식 방법은 검출하고자 하는 객체의 이미지를 레이블링 하고 데이터셋화 한 후 네트워크 아키텍처를 설계하고 모델을 완성시켜 활용한다. 수천 또는 수백만 장의 이미지로 훈련된 AlexNet[3] 또는 GoogLeNet[4]과 같은 기존 네트워크를 사용하여 결과물을 빠르게 산출할 수 있지만 모든 분야에 적용하기는 어렵고 특정 시스템에 적용하기 위해서는 커스터마이징이 필수적이다. 기존의 현장 영상서비스를 위한 데이터셋 커스터마이징 방법은 현장의 영상데이터를 영상 그대로 수집하거나 외부(인터넷 등)에서 관련있는 데이터를 크롤링하여 획득하고 객체 레이블링을 통해 데이터셋을 구성하였다. 이 데이터들을 확충하기 위해 데이터를 어그멘테이션(augmentation)하는 형태가 주였고 그 양에 비례하여 많은 수작업 시간과 노력이 수반되었다. 이미지 어그멘테이션은 기존의 이미지 데이터의 정보량을 유지한 상태로 노이즈를 주는 방식이 대표적이다. 하지만 정보량 자체는 변하지 않기 때문에 데이터 증강 관점에서는 한계

점이 존재한다. 따라서 정보량 자체는 늘리고 방대한 영상데이터를 자동으로 가공하는 효율적인 커스텀 모델을 생성하고 지속적인 전이학습을 통해 모델 성능을 향상해 가야 하며 현장에 맞춤형 모델로 적용이 필요하다.

상황인식 측면에서는 비정형화된 시공간객체(ex. 불, 연기, 안개 등) 및 상황을 인식하는데 있어서 객체 인식 알고리즘으로는 한계점이 존재하고 기대하는 결과를 얻기 위해서는 여러 가지 알고리즘과 분석론이 필요하다. 산업 다방면에서 필요로 하는 분야지만 적용 기법이 다르고 명확한 원인 규명의 어려움, 이상값을 가진 데이터 확보의 어려움 등 여러 가지 한계점이 있지만 최근까지도 활발히 연구가 이뤄지고 있는 분야이다[5].

영상관리 측면에서는 영상 저장, 요약, 연계, 검색, 변환, 클라우드, 스토리지 관리 기능이 구현되어야 한다. 에이전트를 통한 객체 인식 영상분석 프로세스를 기반으로 경고 수준, 사건 발생 수준 등 수준별 이벤트 대응, 장비제어, 이동 객체 동선 분석, 장면별 영상 의미 인덱싱 등의 알고리즘이 각 적용 분야에 맞게 커스터마이징 되어 적용이 가능해야 한다[2].

따라서 영상수집부터 영상관리, 학습데이터 수집부터 학습까지, 영상의 분석부터 상황인식까지 각 컴포넌트들을 비즈니스 요구와 영상서비스의 형태에 맞게 자동화 및 통합하여 시스템화 할 필요성이 있다.

본 논문에서는 한국전력공사의 인프라 및 전력설비, 장비, 드론 영상 및 공공 CCTV의 수집된 영상데이터를 기반으로 영상전환점을 자동으로 추출한 다음 커스텀 학습을 수행하여 현장 영상서비스에 최적화된 모델을 생성하여 비정형객체를 인식하고 이상을 탐지하는 방법을 제공하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문의 기본 이론인 영상전환점 추출, 객체 인식 및 탐지 기술, Class Activation Map[6], Gan[7]에 대하여 정의한 다음, 3장에서 제안하는 시스템을 설명한다. 그리고 4장에서는 제안한 시스템 흐름대로 실험한 결과를 객관적인 지표와 성능평가 트릭으로 성능을 평가하고 5장에서 결론을 맺는다.

2. 이 론

2.1 영상전환점

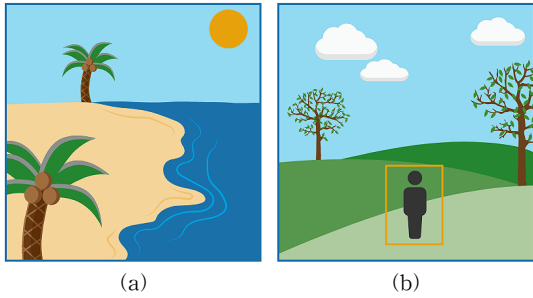


Fig. 1. Computer vision technology. (a) Object recognition and (b) Object detection.

영상전환점은 영상 프레임 사이의 컷을 감지하는 것으로 콘텐츠 인식 감지와 임계값 기반 감지로 나눌 수 있다. 콘텐츠 인식은 두 개의 연속 프레임 간의 차이가 설정된 임계값을 초과하는 영역을 찾는 것으로 임계값을 적절히 설정하면 필름의 점프컷과 같은 사소한 갑작스러운 변화도 감지할 수 있다. 임계값 기반 감지는 현재 프레임의 강도와 밝기를 설정된 임계값과 비교하고 이 값이 임계값을 초과할 때 장면 컷/브레이크를 트리거하여 작동하는 방식이다. 프레임의 모든 픽셀에 대한 RGB 값을 평균화하여 계산되어 평균 픽셀값(0.0~255.0)을 나타내는 단일 부동 소수점 숫자를 산출한다.

2.2 객체 인식 및 탐지

객체 인식이란 비디오 또는 이미지상의 객체를 식별하는 컴퓨터 비전 기술로 객체 인식은 딥러닝과 머신러닝 알고리즘 기반 핵심 기술이다. 사람은 비디

오 또는 사진을 볼 때 사물, 장소, 장면 등 시각적 세부 사항을 쉽게 알아볼 수 있다. 이처럼 사람이라면 쉽게 할 수 있는 일을 컴퓨터가 할 수 있도록 학습시키는 것을 말한다. Fig. 1의 객체 인식(object recognition)과 객체 탐지(object detection)는 유사한 객체 식별 기술이지만 구동 방식은 서로 다르다. 객체 탐지는 이미지에서 객체의 인스턴스를 찾아내는 것으로 딥러닝에서는 이미지에서 객체를 식별할 뿐만 아니라 위치까지 파악할 수 있는 객체 인식의 부분집합이다. 이를 통해 하나의 이미지에서 여러 객체를 식별하고 각 위치 또한 파악할 수 있다[8].

2.3 CAM (Class Activation Map)

CNN(Convolutional Neural Network)[3] 특정 클래스 이미지를 해당 클래스라고 예측하게 한 그 이미지 내의 위치 정보를 의미한다. 기본 구조는 Goog LeNet과 Network in Network와 흡사하며 아래의 Fig. 2에서 볼 수 있듯이 마지막 conv layer를 fc-layer로 flatten 하지 않고, GAP(Global Average Pooling)을 통해 weight들을 생성해 낸다. 마지막 conv layer가 총 n개의 channel로 구성되어 있다면, 각각의 채널들은 GAP을 통해 하나의 weight값으로 나타내고 총 n개의 weight들이 생기게 된다.

최종적으로 softmax 함수로 연결하여 이 weight들도 백프로를 통해 학습을 시킨다. N개의 weight가 생겼다면 이 weight들과 마지막 n개의 conv layer들과 weighted sum을 해준다. 그러면 하나의 특정 클래스 이미지의 heat map이 출력하게 된다. 이 클래스

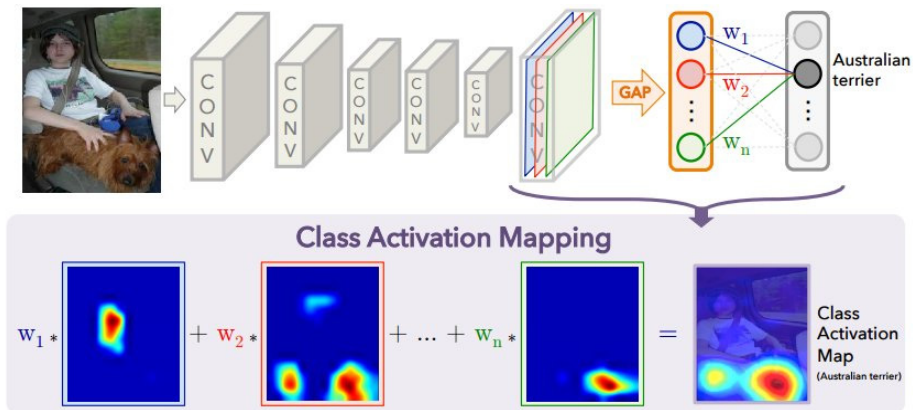


Fig. 2. Class activation map.

는 ‘Australian terrier’ 로 개의 종류중 하나인데 개의 몸 부분과 얼굴부분을 보고 ‘Australian terrier’로 판단했다는 것을 Fig. 2에서 확인할 수 있다.

2.4 GAN (Generative Adversarial Network)

생성적 적대 신경망으로 보통 신경망 모델들과 다르게 생성자와 판별자 2가지 종류의 신경망을 가지고 있으며 진짜 같은 가짜를 생성해내는 것을 목표로 하고 있는 딥러닝 알고리즘이다. 생성적이라는 의미는 데이터를 스스로 생성하는 생성자를 뜻하며 훈련 데이터셋에 따라 결정되기에 피카소의 작품처럼 만들고 싶을 경우 피카소의 작품 데이터셋을 준비하면 된다. 적대적이라는 의미는 생성자와 판별자의 관계를 뜻하는 것이며 생성자가 가짜를 진짜처럼 계속 만들어내는 과정에 있어서 판별자는 점점 더 진짜 같은 가짜를 감별해야 하며 서로 이기고 지는 싸움을 수없이 반복하면서 모델은 더욱 정교해진다. Min-max 게임이라고 해서 하나의 V가 있을 때, 판별자는 커지는 방향으로 업데이트를 하고, 생성자는 반대로 이러한 식이 감소하는 형태로 네트워크를 업데이트 한다. 동일한 함수에 대해서 한 쪽은 값을 낮추고 한 쪽은 값을 높이는 방식으로 업데이트를 진행한다. 수식 (1)에서 x를 sampling해서 $[\log D(x)]$ 라는 식에 대입한다. 또한 z를 sampling 하기 위한 판별자는 일반적으로 gaussian 분포를 사용한다. 마찬가지로 z를 매번 sampling 한 뒤에 $(\log(1-D(G(z))))$ 에 대입하고 기댓값(E) mean값을 구한다. 판별자는 이미지인 x가 들어왔을 때 이때 x는 원본데이터의 분포를 따르므로 원본데이터에 대해서는 1을 도출할 수 있도록 학습을 진행한다. 결론적으로 생성자는 진짜 이미지를 진짜라고 분류하고 판별자는 가짜 이미지를 가짜 이미지로 분류 하도록 학습을 진행한다. 그 과정에서 생성자가 만드는 이미지는 실제 이미지와 비슷하게 보이도록 학습이 되며 학습이 다 진행이 되면 이 판별자는 더 이상 실제 이미지와 가짜 이미지를 구분할 수 없게 되면서 output값은 0.5로 수렴하게 된다. 생성자에서 sampling되는 image들은 이러한 real image distribution을 따를 수 있도록 수렴한다[9].

$$\min_G \max_D V(D, G)$$

$$V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

3. 제안한 방법

3.1 제안 시스템 개요

본 논문에서는 영상전환점 및 객체 탐지를 기반으로 입력 영상의 프레임 변화량을 분석하여 이상을 탐지하고, 실제 영상서비스 운영환경에서 수집한 영상데이터 커스텀 학습을 통해 생성한 모델로 비정형 객체를 인식한다. Fig. 3에 제안하는 시스템의 전체적인 흐름도를 나타내었으며 영상 입력, 영상전환점 추출 및 객체 탐지, 이상 탐지를 위한 Class activation map, 비정형객체 인식, 알람 출력으로 이루어져 있다.

3.2 영상전환점 추출 및 객체 탐지

본 논문에서는 영상전환점 추출을 위해 Pyscene-detect[10] 라이브러리를 활용하였다. 콘텐츠 인식 감지와 임계값 기반 감지를 통해 매프레임의 hue, saturation, luminance를 설정된 threshold값과 비교하고 이 값을 초과할 때 영상전환점을 추출한다. 일반적으로 한 이미지에서 객체의 존재를 감지하려면

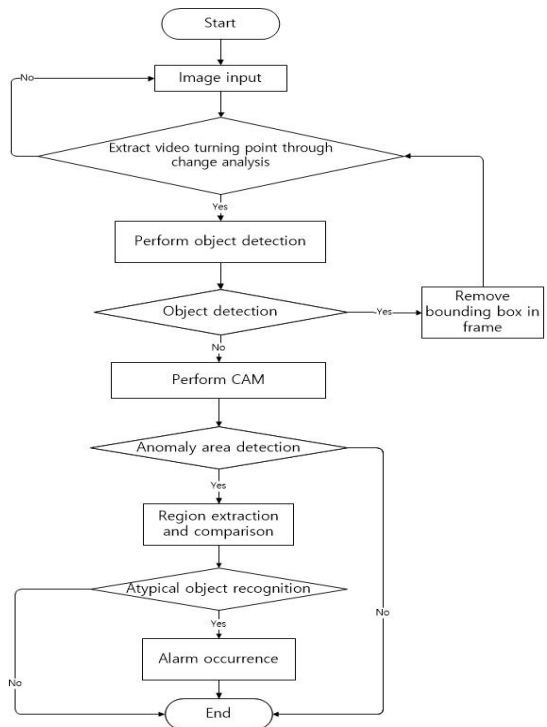


Fig. 3. Proposal system flowchart.

이 객체의 높이가 뷰의 10%를 차지할 수 있어야 하며 알고 있는 객체를 인식하려면 50% 추가적으로 식별하기 위해서는 120% 이상이 필요할 수 있다.

또한 객체의 출현 및 이동을 감지할 때 객체가 뷰가 50~100%이면 threshold의 적절한 값은 10이며 작을수록 민감도가 증가하는 것(미세한 변화에도 영상전환점 추출)을 확인하였다. 예를 들어, 사람의 얼굴은 독특한 식별 특성을 가지고 있으며 얼굴 너비는 신체의 높이나 너비보다 변화가 적다. 사람 얼굴은 평균 폭은 16cm(6.3인치)이며 사람의 얼굴을 객체로 가정했을 때 이러한 특성을 토대로 아래 Table 1을 참고하면 적절한 임계값을 도출할 수 있다[11].

운영 요건의 어려운 조건은 변화가 많고 약한 조명 상태로 사람, 물체, 차량에 대한 모니터링이 일정 각도에서 이루어지지 않고 그늘에 있거나 추가 조명이 없는 실외상황이나 매우 어두운조건의 실내 상황을 말하며 디테일이 그늘에 있거나 카메라를 향하지 않고 있는 경우, 여러 객체가 한 구역을 통과하는 상

황도 포함 된다. 좋은 조건은 조명이 좋은 상황으로 사람, 물체, 차량이 합리적인 속도로 움직이며 모니터링이 일정각도에서 이루어지고 충분한 디테일을 볼 수 있으며 추가 조명이 있는 실외 상황이나 조명이 균일한 실내 상황을 말한다.

실시간으로 영상전환점이 추출되었다면 해당 프레임에 대한 객체 탐지를 수행하고 Fig. 4처럼 객체가 감지되었다면 해당 bounding box를 ROI(Region Of Interest)[12]로 설정한다. 영상전환점 추출 시 frame number와 timecode 값을 기반으로 관심 영역을 실시간으로 추적하여 해당 좌표를 프레임에서 제거 또는 객체가 없는 프레임의 좌표로 교체하고 영상전환점 추출을 지속한다[13]. 이러한 기법은 이상 탐지를 원하지 않는 객체의 출현 및 이동으로 인한 영상전환점 추출을 제어할 수 있다.

추가적으로 영상전환점이 추출되었고 객체가 존재하는 프레임이라면 해당 프레임은 객체 탐지를 위해 학습 가치가 있는 프레임으로 볼 수 있다. 즉 자동

Table 1. Facial identification requirement indicators.

| Operational requirements | Horizontal pixels/face | Pixel/cm | Pixel/inch |
|--------------------------------------|------------------------|-----------|------------|
| Identification (difficult condition) | 80px/face | 5px/cm | 12.5px/in |
| Identification (good condition) | 40px/face | 2.5px/cm | 6.3px/in |
| Recognize | 20px/face | 1.25px/cm | 3.2px/in |
| Detect | 4px/face | 0.25px/cm | 0.6px/in |

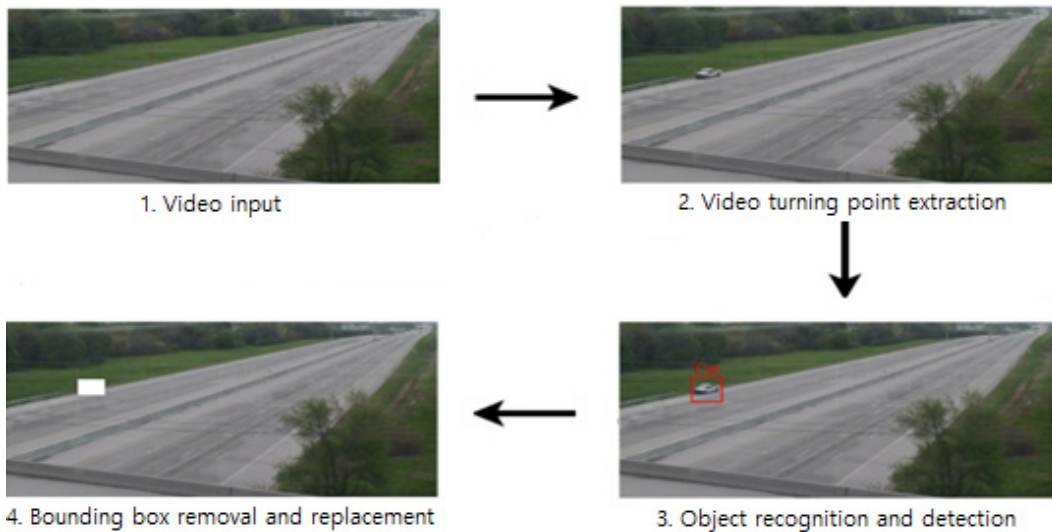


Fig. 4. Video turning point extraction and object detection flow.

으로 레이블링 작업이 요구되는 학습 가치가 있는 이미지들이 수집되는 것이다. 앞서 2.4에서 정의한 Gan을 통해 자동으로 수집되며 이러한 학습 가치가 있는 프레임들은 2D 이미지들을 3D 형상 가상이미지로 증강시킬 수 있다.

또한 Fig. 4에서 bounding box를 제거 했을 때 증강된 가상 이미지들로 교체하고 감지 여부 및 정확도를 수치화하여 GAN의 성능평가 기법인 Frechlet Inception Distance[14]와 다른 실제 검증 방식으로 증강된 이미지들의 성능 평가를 할 수 있다. GAN의 사전조건은 유사한 데이터셋을 구성하는 것인데 이러한 방법으로 현장에 특화된 학습 데이터를 자동

수집 및 증강시키고 평가할 수 있다. 영상서비스 운영 관점에서는 정기적 학습 스케줄러를 등록해 놓는다면 자동으로 자가학습 및 모델 업데이트를 통해 현장 맞춤형 모델을 최적화하여 실시간으로 최신의 영상서비스를 제공할 수 있다.

3.3 이상 탐지 판단을 위한 Class Activation Map

본 논문에서는 객체출현 및 자연재해나 사회재난 발생을 이상 탐지 조건으로 설정하였고 영상전환점 이상 탐지 판단을 위한 class activation map은 Scipy, Numpy, Keras 라이브러리를 활용하여 구현하였다. 수행에 앞서 기본적으로 이상 탐지 객체가 학습된

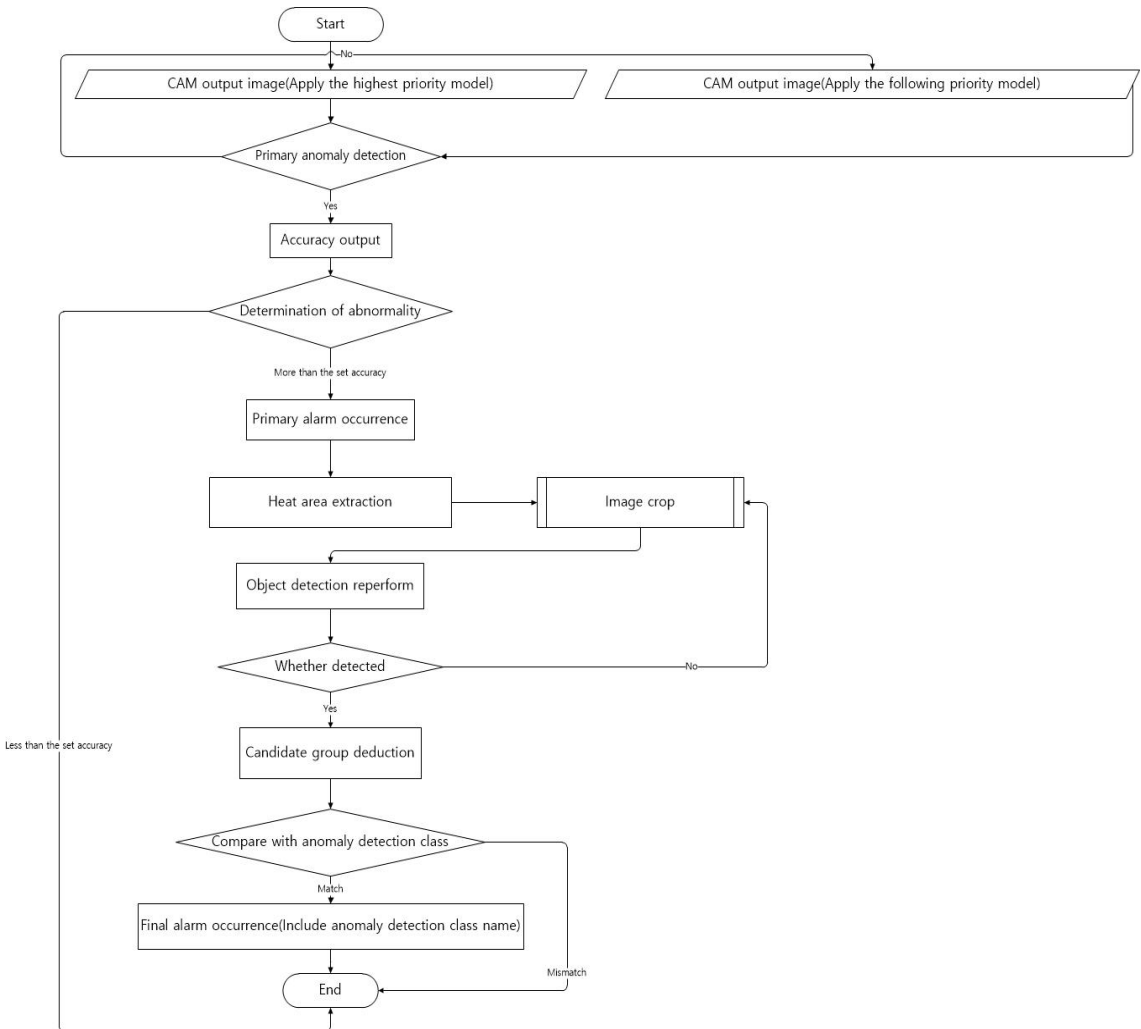


Fig. 5. Unstructured object recognition algorithm.

모델이 필요하고 비정형객체인 불과 연기가 각 각 학습된 2가지 모델을 사용하였다. 모델 데이터셋은 현장의 정상 이미지 클래스와 이상 탐지 객체가 포함된 이미지 클래스로 두 클래스로 구성하였고 Mobile NetV2[15]로 이미 학습되어 있는 모델을 개조하고 전이학습하여 커스터마이징하였다.

현장에 맞춤형으로 적용하기 위해 정상이미지 클래스는 계절과 밤과 낮, 자연현상, 기상이변, 특수상황 등이 고려된 데이터셋을 활용하였고 다른 클래스는 이상 탐지 객체가 포함된 이미지와 인위적으로 이상 탐지 객체를 가공한 이미지가 포함되어 구성되어 있다.

3.4 비정형객체 인식

일반적으로 객체 탐지는 사각 영역기반 객체 탐지 기법들이 주로 사용된다. 하지만 사각영역이 아니면서 매우 큰 범위를 가지거나 형태가 없는 비정형객체 인식에는 적절치 않다는 단점이 있다. 현장 특성에 맞는 학습된 모델로 class activation map을 수행하며 이상이 있다고 판단했다면 Fig. 5의 알고리즘 수행한다. 1차적으로 predict한 결과를 출력한다. 이 결과는 이상 탐지가 객체가 존재할 확률을 출력한다. 설정된 이상 탐지 accuracy 이상이라면 해당 이미지에서 이상 영역(heat 영역)을 정확히 추출하고 image processing하여 분석하는 방법은 실시간 처리에 한계가 있으므로 이미지를 4등분으로 crop한 후 heat 영역과 가장 많이 겹치는 crop된 이미지를 우선순위로 하여 object detection을 재수행한다. 모든 프레임을 crop하고 sliding window 방식으로 object detection을 수행하면 탐지율은 높아지지만 실시간 처리와 연산량 최적화를 위해 이 방법을 사용한다.

비정형객체 특성상 시간에 따라 형태가 바뀌며 object detection으로 객체를 탐지 시 비교적 같은 area에서 탐지되는 성격을 지니고, 한 프레임에서 동일한 여러 bounding box가 탐지될 수 있으며 연속된 프레임에서도 마찬가지이다. 이러한 규칙들은 분석하여 적절히 적용하면 보다 나은 판단 결과를 도출할 수 있다. 탐지된 bounding box에 index값을 매겨주어 배열의 모양들을 보고 6개의 파라미터(대각선 길이, Theta의 최대값, 면적의 차이, 너비 차이, 높이 차이, 최소 그룹)를 사용하여 재귀함수로 후보군을 도출한다. 즉, 한 프레임에서 근거리에 같은 클래스

의 bounding box가 한 개 이상 탐지될 경우 후보군 우선순위가 높아진다. 최종후보군이 도출되면 1차 predict 결과와 비교하고 비정형객체가 무엇인지 최종 판단하고 알람을 출력한다.

4. 실험 결과 및 고찰

제안된 시스템의 성능을 검증하기 위해서 영상전환점 추출에는 실제 사고사례로 보도되었던 연합뉴스의 실제 산불 발화 추정 영상을 사용하였고, Class activation map은 유형별 성능 검증을 위해 영상전환점 추출 시 사용했던 영상과 공사 현장 샘플 이미지, NASA Space Apps Challenge의 fire dataset[16]을 사용하였다.

Fire dataset은 2개의 클래스로 나누어지며 화재 이미지 755장과 비화재 이미지 244개의 이미지로 이루어져 있다. 모델 성능 평가 방법으로는 분류 정확도(Accuracy)를 사용하였다. 학습 데이터는 훈련 데이터 90%, 검증 데이터는 10%로 설정하여 10번 반복하였으며 학습 결과 best 모델 정확도는 99.11을 기록하였다.

Table 2. Video turning point extraction result.

| Threshold/ value | Frame/ frames | Detected object/objects | Whether ignition point is detected |
|---------------------|------------------|----------------------------|--|
| 1 | 191 | 14 | ○ |
| 5 | 81 | 4 | ○ |
| 10 | 26 | 4 | ○ |
| 20 | 7 | 2 | ○ |
| 30 | 6 | 1 | ○ |



Fig. 6. First ignition point frame. A spark occurs in the upper left corner of the frame at the time of the first fire.

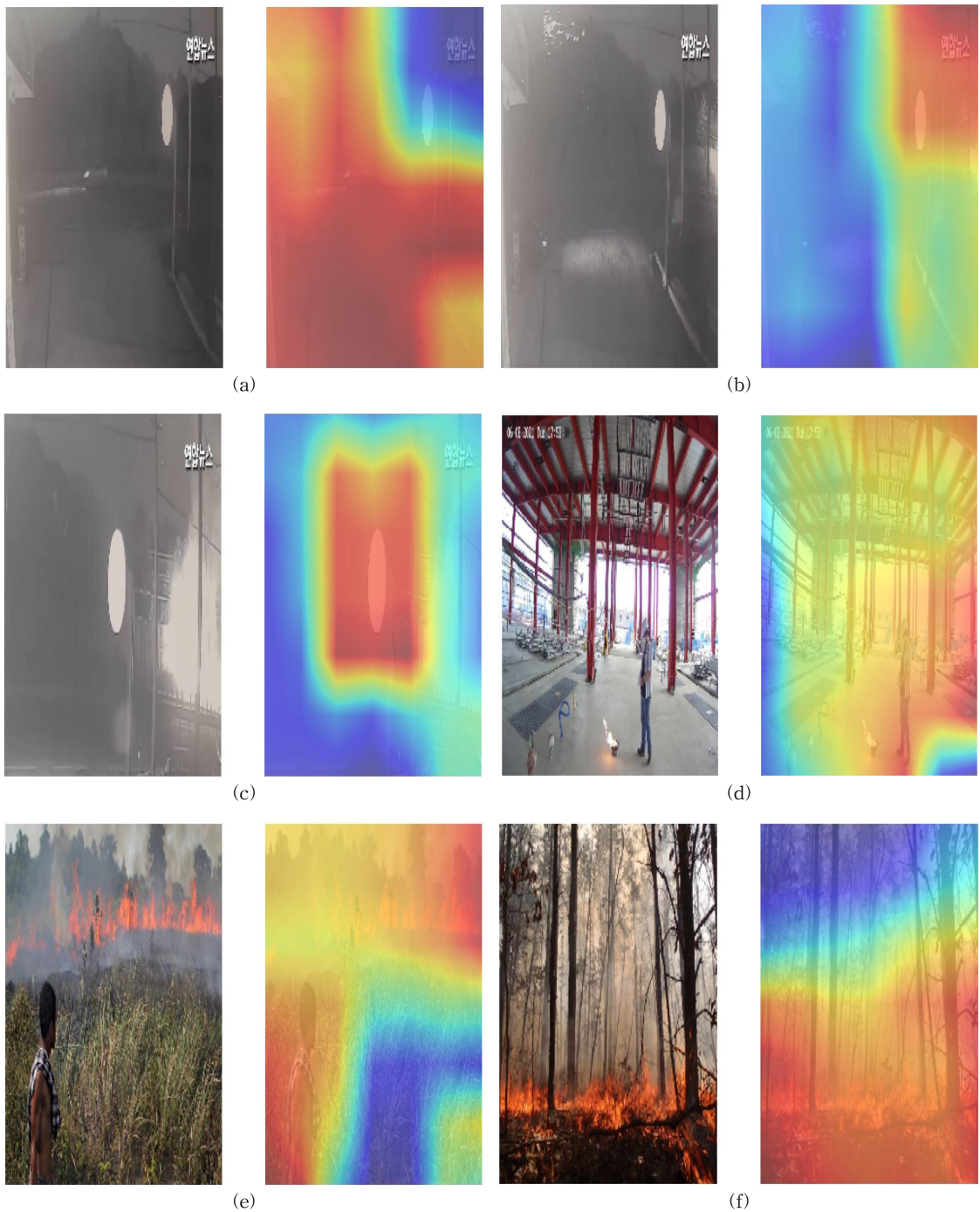


Fig. 7. Prediction results by running the class activation map, (a) Normal frame input result, Probability of presence of fire : 12.5%, (b) First frame with sparks, Probability of presence of fire : 99.06%, (c) Frame on fire, Probability of presence of fire : 88.44%, (d) Fire frame on construction site, Probability of presence of fire : 100%, (e) The probability of no wildfire is 2.05%, That is, the probability that a wildfire exists within the frame : 97.95%, (f) The probability of no wildfire is 1.46%, That is, the probability that a wildfire exists within the frame : 98.54%, (g) The probability of no wildfire is 100%, and (h) The probability of no wildfire is 100%.



Fig. 7. Continued.



Fig. 8. Results of performing unstructured object recognition, (a), (c) A screen in which unstructured objects are not detected when performing the first object detection and (b), (d) A screen with object detection through video integration analysis.

4.1 영상전환점 추출 결과

영상수집 약조건의 시나리오로 야간 상황에서 해상도가 1280×720인 3,104장의 프레임(약 1 분 영상)을 threshold값을 변화하여 영상전환점 추출 결과는 아래의 Table 2와 같다. 육안검사 시 객체 출현 개수는 14개였으며 아래 Fig. 6의 최초 스파크 발생 프레임의 경우 모든 threshold 구간에서 추출되었다.

또한 동일한 영상을 객체 탐지 알고리즘중 성능이 우수한 Yolov5[17]로 가장 정확도가 높은 모델인 yolov5x 모델을 통해 객체 탐지를 수행했을 때 객체가 거의 검출되지 않았으며 탐지된 객체들도 오탐이 많았다. 이는 영상분석 및 상황인식에 있어서 영상수집 환경에 덜 민감하고 미감지가 매우 희소한 영상전환점 추출 기법이 필요함을 시사한다.

4.2 Class activation map 수행 결과

Class activation map은 앞서 4.1에서 추출된 영상

전환점, 실제 공사 현장, fire dataset 순서로 수행하였으며 결과는 아래의 Fig. 7과 같다. (a), (b), (c), (d)는 학습 데이터가 실제 현장 수집데이터의 화면이 아닌 경우로 커스텀데이터와 외부데이터, 학습데이터 종속성 등을 비교하기 위해 실험하였으며 class activation map이 프레임 내 어느 부분을 heat하여 판단하는지 알 수 있도록 시각화하였다.

(e), (f), (g), (h)는 실제 학습에 사용된 데이터로 실제 현장데이터로 학습을 수행할 때 이상 탐지 판단 확률 정확도가 매우 높은 것을 확인하였다. 또한 이상 탐지시 앞 프레임이 다음 프레임에 잔광 영향을 줄 수 있다는 점과 단순 객체 크기에 의해 성능이 종속되지 않는다는 점을 확인할 수 있었다.

4.3 비정형객체 인식 수행 결과

아래 Fig. 8의 (a)는 Fig. 6의 (d)와 동일한 사진이며 yolov5를 객체 탐지 엔진으로 사용하였다. 불꽃부

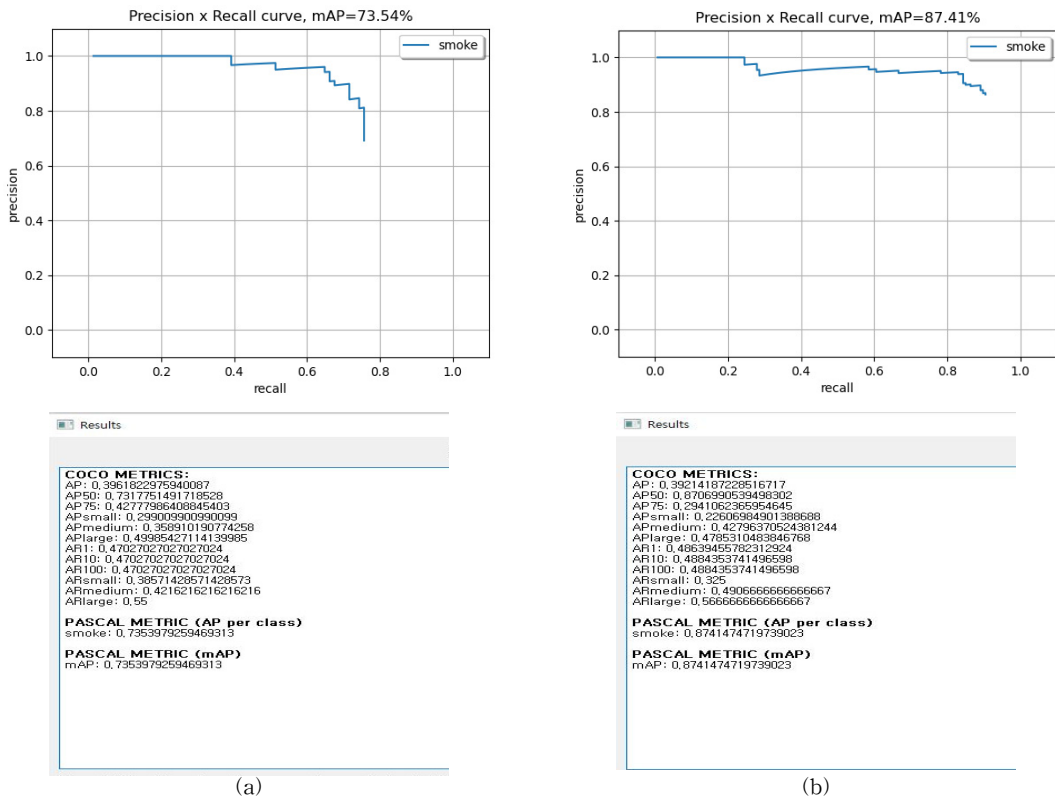


Fig. 9. Performance evaluation metric results. (a) Indicators that do not include unstructured object recognition and (b) Indicators including unstructured object recognition,

터 산물까지 포괄한 훈련데이터 402장, 검증데이터 90장으로 옛지에 최적화된 모델인 yolov5s 모델을 가중치로 하여 학습한 모델(mAP 90% 이상)로 객체 탐지를 수행한 결과이며 jetson tx2 보드로 옛지단에서 구동 및 실험하였다. (c), (d)는 연기 데이터셋 663장을 별도로 학습하여 비정형객체 인식을 수행한 결과이다. (c)는 최초 객체 탐지시 결과이며, (d)는 연기 비정형객체 인식을 수행한 결과이다.

최초 객체 탐지 수행 시에는 화재와 연기가 감지되지 않았었지만 추출된 영상전환점에서 class activation map을 수행하고 이상영역이라고 판단했던 heat 영역과 프레임의 hue, saturation, luminance 변화량을 분석한 영역을 원본에서 crop하여 재탐지한 결과 0.39 confidence로 화재를 최종 탐지하였고 0.49, 0.34, 0.32 confidence로 연기를 최종 탐지하였으며 비정형객체 인식 알고리즘에 의해 최종적으로 연기 이상 탐지 판단 알람을 출력하였다. 이는 지능형 영상 분석을 통합 수행하여 검증된 결과임을 명시한다.

또한 Fig. 9은 비정형객체 중 연기를 ground truth와 detect 결과를 기반으로 성능평가 메트릭을 사용하여 최종 객체 탐지 성능을 측정 한 지표이다. (a)는 비정형객체가 인식 결과가 포함되지 않은 지표이고 (b)는 비정형객체 인식을 포함한 지표이며 mAP가 13.87% 상승한 것을 확인하였다.

5. 결 론

본 논문에서는 영상전환점 및 객체 탐지를 기반으로 비정형객체 인식 및 이상 탐지를 할 수 있는 영상통합분석 시스템을 제안하였다. 제공하고자 하는 영상서비스에 따라 현장 맞춤형 영상전환점 추출 기법과 학습데이터 수집을 위한 자동화된 방법을 제안하였고, 이를 학습하기 위한 전이학습과 자가학습 구조를 제안하였다. 실험 결과, 영상전환점 추출 기법을 통한 이상 탐지는 현장 약조건에서도 100% 영상전환점을 추출하는 우수한 성능을 보였고, class activation map은 적은 데이터셋임에도 각 이상 탐지에 특화된 학습 모델에 따라 약 13.87% 성능 상승 변화 효과성을 실험을 통해 입증하였다. 이는 제안된 비정형객체 인식을 위한 영상통합분석시스템이 이상 탐지 문제 해결에 도움이 될 수 있음을 시사한다. 본 논문에서는 자연재해 및 사회재난을 이상 탐지로 비록 작은 범위로 설정하였지만 사회 다방면에서 제안

한 통합 시스템이 영상서비스 분야에 목적 지향적인 실제적인 통찰력을 제공할 수 있으며 전환점이 될 것으로 기대한다.

REFERENCE

- [1] NIPA, *Understanding and Utilization of Visual Intelligence*, Issue Report 2019-09, p. 2, 2019.
- [2] S.-Y. Ok, "Real-Time Large-Capacity/ Large-Scale Video Data Distributed Agent-Based for Analysis to Develop a High- Performance Object Tracking Platform," *A Study on Korea (Research)*, pp. 3-7, 2017.
- [3] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network," *NIPS*, pp. 1097-1105, 2012.
- [4] C. Szegedy, et al. "Going Deeper with Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.
- [5] Anomaly Detection, http://docs.iris.tools/manual/IRIS-Usecase/AnomalyDetection/AnomalyDetection_202009_v01.html (accessed October 1, 2021).
- [6] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2929, 2016.
- [7] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, and D. Warde-Farley, *Generative Adversarial Networks*, *arXiv Preprint*, arXiv: 1406.2661, 2014.
- [8] Computer vision technology, <http://www.aitimes.kr/news/articleView.html?idxno=12087> (accessed July 24, 2021).
- [9] GAN review, <https://airsbigdata.tistory.com/217> (accessed July 24, 2021).
- [10] Pyscenedetect, <https://pyscenedetect.readthedocs.io/en/latest/other/literature/> (accessed July 24, 2021).
- [11] Axis, <https://www.axis.com/ko-kr/learning/>

web-articles/perfect-pixel-count/pixel-density (accessed June 20, 2021).

[12] H.S. Parekh, D. Thakore, and U.K. Jaliya, "A Survey on Object Detection and Tracking Methods," *IJIRCCE*, Vol. 2, pp. 2970-2978, 2014.

[13] S.-C. Lim and J.-C. Kim, "Bottleneck-based Siam-CNN Algorithm for Object Tracking," *Journal of Korea Multimedia Society*, Vol. 5, No. 1, pp. 72-81, 2022.

[14] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*, *NIPS*, Vol. 30, pp. 6629-6640, 2018.

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018.

[16] Fire dataset, <https://www.kaggle.com/phy-lake1337/fire-dataset> (accessed September 12, 2021).

[17] Yolov5, <https://github.com/ultralytics/yolov5>, 202 (accessed April 15, 2021).



이 석 창

2013년 한국공학대학교 컴퓨터공학과(공학사)
 2013년~2015년 한국저작권위원회 SW포렌식팀
 2019년~2020년 한국수력원자력(주) 기계연구소

2021년 KAIST 지식재산공학(공학석사)
 2021년~현재 KAIST 소프트웨어 전문가 통합과정
 2020년~현재 한국전력공사 전력연구원 디지털솔루션연구소
 관심분야: Visual Intelligence, Multimedia Design, Hyperautomation, Industrial Safety AI



김 영 현

2002년 한국항공대학교 통신정보공학(공학사)
 2004년 광주과학기술원 정보통신공학(공학석사)
 2004년~현재 한국전력공사 전력연구원 책임연구원

관심분야: Communication System, Wired and Wireless Communication Network, IoT



강 수 경

2008년 한양대학교 전자전기컴퓨터공학부(공학사)
 2008년~현재 한국전력공사 전력연구원 선임연구원
 관심분야: A.I, Smart Grid, Image/Video Analysis



박 명 혜

1993년 경북대학교 전자공학과(공학사)
 1995년 경북대학교 전자공학과(공학석사)
 1995년~현재 한국전력공사 전력연구원 수석연구원

관심분야: IoT, Wireless Communication, Intelligent Video Analysis