https://doi.org/10.7840/kics.2023.48.8.1001

특정 객체 필터링 및 이상동작 감지를 위한 통합 합성곱 신경망

홍 상 욱◆, 김 형 원°

Integrated CNN for Specified Object Filtering and Abnormal Motion Detection for Smart Factory

Sang-wook Hong◆, Hyung-won Kim°

요 약

이상 감지를 위한 합성곱 네트워크 모델로서 오토인코더[1,2,3] 네트워크가 많이 사용되고 있다. 오토인코더는 입력 이미지를 기존에 학습한 이미지 데이터와 유사한 데이터로만 국한하여 출력 이미지를 재생성 하는 특성을 가지고 있어서 입력 이미지 내의 이상 동작을 추출할 수 있는 장점이 있다[4]. 그러나 이러한 방법은 이미지내에 정상적 객체나 상황이 불규칙적인 위치에 포함된 경우는 정상 또는 이상 동작으로 정확하게 감지하기 매우 어려운 단점이 있다. 이 문제를 개선하기 위해 오토인코더와 객체 분할 헤드를 동시에 가진[5] 통합 합성곱 신경망을 제안한다. 본 제안 기술은 입력 이미지와 재생성 된 출력 이미지 내에서 정상 객체 영역을 제외하기 위한 객체 마스킹을 통해 해당 객체가 불규칙적인 위치에 포함된 이미지에서도 정확도를 증가시킨다. 또한, 제안 기술은 통합 합성곱 신경망에 포함된오토인토더를 이용하여 마스킹 대상인 객체들의 라벨을 자동으로 생성하고[6] 객체 분할 모델을 학습시키는 비지도[7] 학습방법을 제안한다. 제안 통합 합성곱 신경망 모델을 스마트 팩토리에서 획득한 이상 동작 비디오 데이터셋에 적용하여 성능을 분석하였다. 통합 합성곱 신경망 모델에 포함된 객체 분할 모델의 마스크 mAP 비교 결과, 제안하는 의사 라벨링 기반의 비지도 학습 기술은 수작업으로 라벨 된 정답 데이터로 테스트할 경우 96.82%의 마스크 mAP의 매우 높은 성능을 보인다. 또한 제안하는 통합 합성곱 신경망 모델은 정상과 비정상 객체가 공존하는 상황에서 기존의 오토인코더가 탐지하지 못하는 정상 객체에 대한 정확도를 증가시켜 전체적인 이상감지 정확도를 15.90% 향상시킨다.

키워드 : 스마트 팩토리, 이상감지, 딥러닝
Key Words : smart factory, anomaly detection, deep learning

ABSTRACT

Autoencoders[1,2,3], which reconstruct the image at the output only to an image similar to their training

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2022R1A5A8026986, RLRC), 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-01304, Development of Self-learnable Mobile Recursive Neural Network Processor Technology, 차세대지능형반도체기술개발사업), 또한 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음" (IITP-2023-2020-0-01462, Grand ICT연구센터), 그리고 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1F1A1061314, 연구재단 기본연구).
◆ First Author : ChungbukUniversity Department of Electrical·Electronic·Information·Computer Science, swhong@msislab.com, 정회원
° Corresponding Author : Chungbuk University Department of Electrical·Electronic·Information·Computer Science, hwkim@chungbuk.ac.kr, 종신회원
논문번호 : 202212-305-C-RN, Received December 20, 2022; Revised February 12, 2023; Accepted June 7, 2023

images, have been widely used in anomaly detection networks[4]. They has the advantage of extracting abnormal motions in the input image because it has the characteristic of regenerating the output image by limiting the input image to data similar to the previously learned image data[4]. However, the anomaly detection method based on an autoencoder has a disadvantage in that it cannot correctly detect a situation in which there is an object to be judged normal at an irregular location in the input image. To address this problem, we propose an integrated CNN with both an autoencoder and an object segmentation heads[5]. We introduce a segmentation-based object masking technique that can exclude normal areas of the input image and reconstructed output image. It can improve the accuracy of detecting anomalies in images where the normal objects are positioned in an irregular location in the image. In addition, we propose an automatic labeling technique which utilizes an autoencoder of the integrated CNN to add pseudo labels[6] to the images containing normal objects and so effectively conducts unsupervised learning[7]. We applied the proposed integrated CNN model to a video dataset obtained from a smart factory. Our unsupervised learning technique demonstrated a mask mAP of 96.82% for the segmentation model when tested using hand-labeled ground-truth data. In addition, the proposed integrated CNN model improves the overall anomaly detection accuracy by 15.90% by increasing the accuracy of detecting normal objects under the situation where both normal and abnormal objects coexist, which the existing autoencoder was not able to detect.

Ⅰ. 서 론

1.1 스마트 팩토리

최근 제조 및 유통을 포함한 생산 과정에 자동화를 적용하는 스마트 팩토리의 도입이 증가하는 추세이며, 따라서 이를 구현하기 위한 데이터 수집 및 그를 위한 이상감지 기술의 중요성이 증가하고 있다.[8] 산업현장에서 강력한 컴퓨팅 파워를 가진 대형 시스템과 고속의 네트워크의 사용이 제한되기 때문에 해당 기술은 소형 사물인터넷 디바이스에서 구동이 가능하며 학습을 위한 데이터의 수집이 간편할 것이 요구된다.[9]

1.2 이상감지[4]

이상감지는 지도학습[11]을 통해 입력의 이상 여부를 분류[12]하거나 이상이 있는 부분을 탐지[13] 하는 기술을 주로 포함하며 최근에 심층 신경망 네트워크를 폭넓게 활용하고 있다. 대부분의 기존 논문들은 객체 분류 또는 검출 합성곱 신경망 네트워크를 사용하며, 학습을 위해서 정상, 비정상여부 또는 비정상 위치를 경계 상자나 마스크로 라벨이 추가된 학습 데이터를 사용한다. 이러한 방법은 네트워크가 이상상황을 학습하기 때문에 다양한 정상상황에서 이상이 발생한 부분만을 상대적으로 높은 정확도로 찾을 수 있다는 장점이 있다. 하지만 훈련을 위해서는 수작업 라벨링이 되어있는 대규모의 데이터셋을 요구하는 제한점이 존재한다.[7]

1.3 비지도[8] 이상감지

공장의 설비 등서 이상감지를 위한 데이터를 수집하는 경우 대부분의 이미지는 정상 동작 이미지들이며, 이상상황을 포함한 이미지들은 전체 데이터에서 극히 일부분만을 차지한다. 일반적인 합성곱 신경망 모델을 지도 학습할 경우 클래스 간의 데이터 비율은 큰 중요성을 가지고 있기 때문에[15] 한가지 분류에 편향된 데이터를 학습할 경우 해당 분류에 과적합이 일어나 모든 입력을 편향된 분류로 판단하는 문제가 발생한다. 이러한 문제를 극복하기 위해 정상 데이터만을 사용해 학습하는 비지도 이상감지[4]가 발표되었다.

일반적으로 비지도 이상감지는 그림 1과 같이 입력과 출력이 같아 훈련 데이터에 라벨링이 요구되지 않고, 입력과 출력의 차인 잔차 오류 이미지를 통해 이상여부를 판별하는 오토인코더[1-3]를 사용한다.

그림 2는 학습 정도에 따른 오토인코더의 입출력 차이를 보여준다. 출력이 입력에 근접할수록 손실 지도에 표시되는 물체가 적어지는 것을 볼 수 있다. 기존 연구들은 오토인코더의 성능을 개선하는 방법에 집중하고

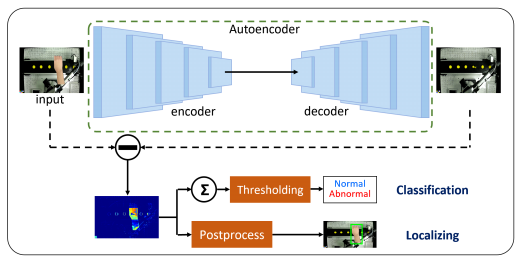
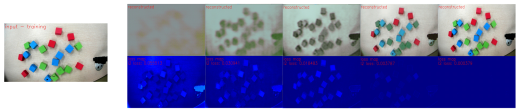


그림 1. 이상감지 오토인코더
Fig. 1. Autoencoder



Input Output and loss map

그림 2. 오토인코더의 학습 정도에 따른 출력 과 픽셀 별 손실
Fig. 2. Output image and loss per pixel according to training degree

있다. 입력을 확률적으로 복구해 정확도를 높이는 Variational Autoencoder Anomaly Detection[16,17], 메모리 계층의 추가로 비정상 입력의 복구를 줄이는 Memory-augmented Autoencoder[18], 입력의 잡음을 격리하여 강인성을 증가시킨 Robust Deep Autoencoder[19] 등 오토인코더 자체의 성능을 높이는 방법에 집중하고 있다. 이러한 기존 기술들은 예측 가능한 정상 객체만 포함하는 영상의 경우에는 이상 동작 검출 정확도를 개선하는 결과를 보인다.

그러나, 반면에 기존의 이상감지 오토인코더는 예측 불가능한 위치에 등장하는 정상 객체를 포함한 영상에 대해서는 이상 동작 검출이 매우 어려우며 또한 이러한 영상에 대해서는 학습 방법에 대한 연구 보고된 바가 없다.

1.4 다중 작업 이상감지

다중 작업이상감지 모델은 다양한 목적으로 사용되고 있다. 일반적으로 Anomaly detection with multiple-hypotheses predictions[20]에서와 같이 기존에 학습한 지식에 대한 정확도를 유지하는 동시에 다양한 도메인을 학습하는 목적으로 사용한다. 본 논문에도 위에서 설명한 다중 작업

이상감지가 가지는 장점을 사용해 네트워크가 정상 배경과 정상 객체 양쪽 지식을 정확도를 유지하면서 학습하는 것을 구현한다.

1.5 특정 객체 필터링 및 이상동작 감지를 위한 통합 합성곱 신경망

기존 지도학습이 가진 제한점인 학습 데이터 수집의 어려움과, 비지도 학습이 가진 예측할 수 없는 위치에 등장하는 정상 객체 학습의 제한점을 극복하고자 본 논문에서는 오토인코더를 이용하는 분할 마스크 라벨링 자동화 기법과 분할[21,22] 네트워크를 이용한 정상객체 마스킹 단계를 추가하여 이상감지 오토인코더의 정확도를 높이는 기법을 제안한다. 또한 정상객체 마스킹 기법을 추가함으로써 증가하는 파라메터 수를 줄이기 위해 오토인코더의 인코더와 분할 네트워크의 백본 네트워크를 서로 공유하는[5] 다중 작업 통합 합성곱 신경망을 제안한다.

Ⅱ. 특정 객체 필터링 및 이상동작 감지를 위한 통합 합성곱 신경망

2.1 제안하는 통합 모델의 구조

본 논문에서 제안하는 모델은 스마트 팩토리의 공정 장비에 설치 가능한 임베디드 인공지능 또는 엣지 인공지능 모듈에서 구동하는 것을 목표로 한다. 따라서, 임베디드 환경에 적합하도록 추론속도가 빠르고 파라메터 최소화가 적용된 모델인 YolactEdge[그림 3][23,24]를 기본 분할 구조로 이용하여 통합 합성곱 신경망 모델을

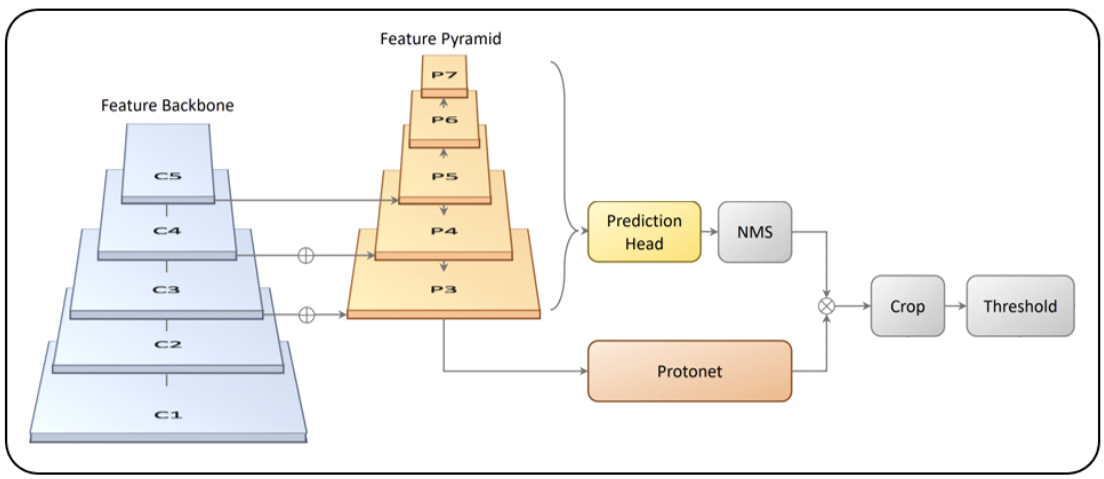


그림 3. YOLACT[23,24]
Fig. 3. YOLACT[23,24]

표 1. 두 모델의 병렬 사용과 다중 헤드 모델의 파라메터 수
Table 1. Number of parameters in parallel use of two models and multihead model

	Params	Params backbone	Params head
Res50 Autoencoder + Yolact edge	60M	47M	13M
Ours	**36M**	**24M**	13M

개발하였다. 서로 다른 두가지모델을 병렬로 구동할 경우 총 파라메터 수가 크게 증가하고 속도가 저하되는 문제가 발생하여 임베디드 인공지능에 적합하지 않다. 따라서 파라메터 수의 증가를 완화하기 위해 두 가지 모델의 계산 복잡성이 높은 네트워크 부분 추출하여 공유가능한 백본 네트워크로 대체하고, 이 백본 네트워크의 출력에 다수의 헤드를 추가하여 다중 작업 합성곱 신경망 구조를 설계했다. 이로 인한 파라메터 수 감소를 표 1에서 확인할 수 있다.

그림 4는 제안하는 다중 헤드 통합 합성곱 신경망 모델의 구조를 나타낸다. 모델은 크게 백본, 특징 피라미드[25], 다중 헤드의 3가지 부분으로 구성되어 있다. 본 논문의 실험에서는 다양한 백본 네트워크를 비교 평가한 후 성능 대비 파라메터 수의 비율이 비교적 높은 레스넷[26]50를 선택하였다. 동작의 순서는 모델에 입력 이미지가 들어오면 레스넷50 기반의 백본이 다양한 크기의 피라미드 단계로 인코딩하여 특징들을 생성하게 된다. 백본의 블록 4가 생성한 특징 3을 다운샘플링 하여 특징4를 획득한다. 특징 3의 업샘플링 결과에 블록 3의 출력을 더해준다. 앞의 과정을 한번 반복해 특징 1, 5를 획득한다. 이렇게 생성한 5개의 특징으로 이루어진 특징 피라미드는 각각의 헤드로 입력되어 분할 마스크와 잔차 오류 이미지를 생성한다. 각 헤드의 출력을 후처리를 통해 통합한다.

상기 특징 피라미드의 디코더 헤드 결과 이미지에 분할 헤드의 결과인 정상객체 마스크로 마스킹 한 후에 잔차 오류 이미지를 출력한다. 이를 통해 정상객체가 존재하는 영상의 경우 거짓 양성을 제거 또는 크게 감소시켜서 전체적인 정확도를 증가시킨다.

2.2 Resnet을 인코더로 사용하는 오토인코더

일반적인 오토인코더는 서로 대칭 구조인 인코더와 디코더를 직렬로 연결하여 순차적으로 수행하는 구조를 가진다. 일반적인 인코더의 레이어 구조는 그림 5의 인코더와 같다. 그러나, 본 논문의 오토인코더는 라벨 자동 생성 기능을 위해 강한 잡음인 정상으로 판단할 객체가 포함된 입력을 받는다. 이를 복구하기 위해 모델은 잡음에 강한, 즉 높은 강인성이 요구된다. 또한 레스넷에 포함되어 있는 잔차 연결이 입 력을 출력으로 복사할 가능성이 있다.

이를 방지하기 위해 모델에 추가적인 일반화 능력이

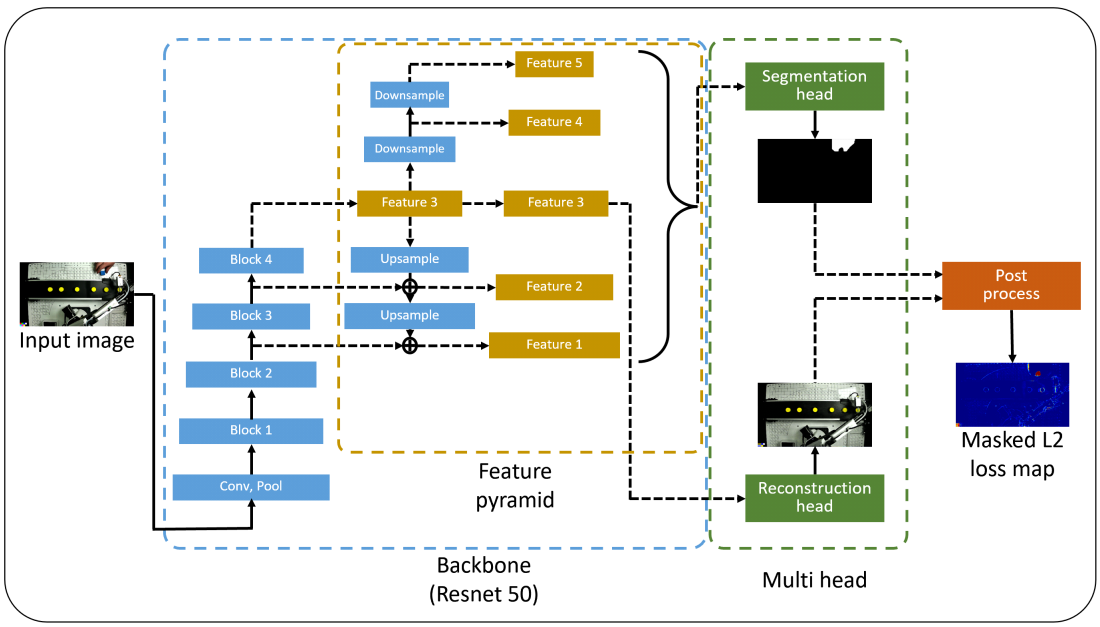


그림 4. 스마트 팩토리를 위한 CNN 통합 객체 필터링 및 이상 동작 감지
Fig. 4. CNN Integrated object filtering and abnormal motion detection for smart factories

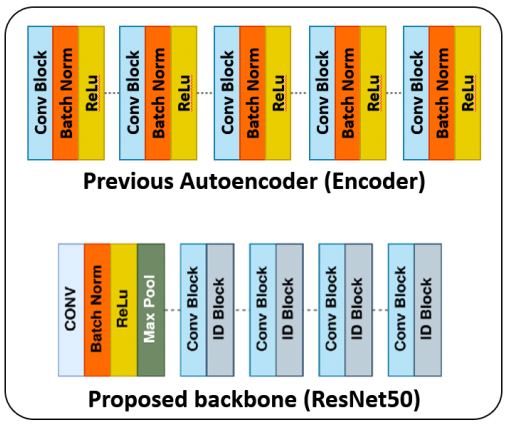


그림 5. 백본 네트워크
Fig. 5. Backbone network

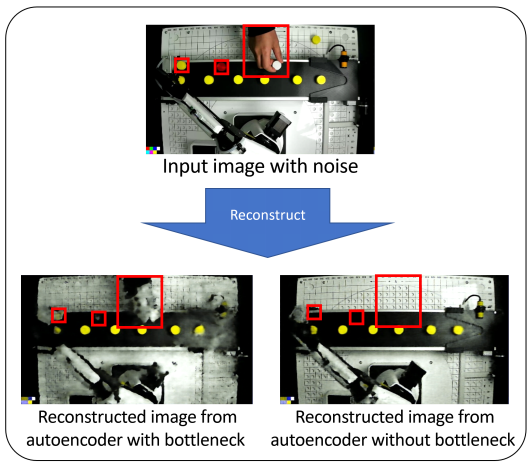


그림 6. 병목 유무에 따른 레스넷 오토인코더의 출력
Fig. 6. Output of ResNet autoencoder with and without bottleneck

표 2. 레스넷을 인코더로 이용한 오토인코더의 제곱 오차
Table 2. L2 loss of autoencoder using Resnet as encoder

	Avg loss under normal video	Avg loss under abnormal video	Difference between Normal and abnormal	Params
Autoencoder	0.00437	0.00673	0.00236	12M
Resnet 34	0.00495	0.00723	0.00228	23M
Resnet 50	**0.00421**	0.00668	0.00247	24M
Resnet 50 freezed	0.00484	**0.00744**	**0.00260**	**24M**

요구된다. 이러한 조건을 만족시키기 위해 디코더 헤드에 입력 이미지의 메모리 크기 7.2Mb의 0.02% 수준인 2Kb를 가진 완전 연결 계층 병목[27]을 추가했다.

그림 6과 같이 병목이 있는 모델은 강한 잡음에도 불구하고 입력의 복구를 원활하게 수행한다. 이에 반해 병목이 없는 모델은 학습하지 않은 입력을 출력으로 복사하는 모습을 보인다. 레스넷 오토인코더의 성능을 검증하기 위해 인코더를 레스넷으로 대체한 레스넷 오토인코더를 구현했다. 정상 이미지와 데이터 증강으로 제작한 비정상 이미지로 테스트를 진행하였다.

레스넷 오토인코더는 표 2에서 이상감지에서 정상, 비정상을 구분하는 지표인 손실 차이가 일반적인 구조의 오토인코더와 거의 유사한 결과를 보인다.

2.3 오토인코더를 이용한 마스크 라벨 자동 생성 기법

그림 7은 라벨 자동 생성의 전체 순서를 보여준다. 제안한 모델의 백본과 비지도 학습으로 훈련시킨 오토인코더 헤드를 이용해 의사 라벨을 생성한다. 오토인코더의 이상여부 판단은 입력과 출력의 차이인 평균 제곱 오차 손실 (L2 loss)[28]를 통해 판별할 수 있다.

$$L2\_loss=\Sigma\ (Input\ -\ Reconstruct)^2 \quad (1)$$

이때 입력과 복구된 이미지의 차이를 시각화 하거나 이상동작을 분할하기 위해 식(1)의 Σ를 생략할 수 있다. 본 기법에선 라벨을 생성할 정상객체를 정확하게 분할할 목적으로 수식(2)를 사용한다.

$$Residual\_error\_image=(Input–Reconstruct)^2 \quad (2)$$

수식(2)로부터 획득한 잔차 에러 이미지에 이진 임계를 적용해 이진 오류 마스크를 얻는다.

$$Binary\_Mask=(Input–Reconstruct)^2>Threshold \quad (3)$$

분할 헤드를 훈련시키기 위해 이 마스크로부터 영역분리 알고리즘을 사용해 각 객체의 영역을 분리해 지역 마스크를 생성한다. Connected Component Labeling, Connected Component Analysis, Blob Extraction[29-31] 등의 이진 이미지에서 픽셀이 집합해 있는 영역을 분리할 수 있는 알고리즘을 사용한다. 영역분리 알고리즘을 적용하기 전 노이즈 제거를 사용할 경우 하나의 영역이 여러 개로 분리되는 문제가 있어 정확도가 감소하는 문제가 있었다. 하지만 영역분리 알고리즘을 적용한 후 노이즈 제거를 실행했을 시 표. 3에서 정답 대비 96.82%의 성능을 보여 후처리에 노이즈 제거를 영영

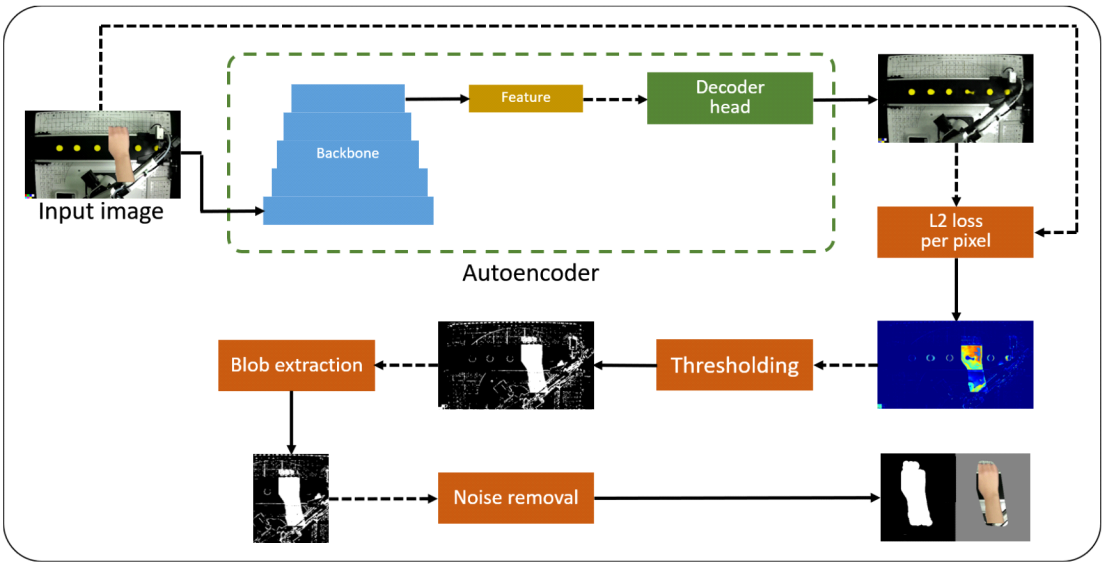


그림 7. 오토인코더를 통한 의사 라벨 생성
Fig. 7. Generate pseudo labels via autoencoders

표 3. 4개의 다른 데이터 세트로 훈련된 제안된 모델의 객체 분할 마스크 mAP 비교: (1) 정답, (2) 노이즈 제거가 없는 의사 라벨, (3) 전체 마스크를 사용하여 노이즈 제거가 있는 의사 라벨, (4) 지역 마스크를 사용하여 노이즈 제거가 없는 의사 라벨
Table 3. mAP comparison of the proposed model trained with 4 different cases: (1) Ground truth, (2) Pseudo Labels without noise removal, (3) Pseudo Labels with noise removal before blob extraction, (4) Pseudo Labels with noise removal after blob extraction

	mAP .50	mAP .55	mAP .60	mAP.50 versus ground truth
Ground truth	97.42	97.37	95.34	100
Pseudo labels without noise removal	91.49	91.30	90.59	93.91
Pseudo labels with noise removal before blob extraction	89.24	88.96	87.92	91.60
Pseudo labels with noise removal after blob extraction	**94.32**	**93.17**	**91.99**	**96.82**

분리 후 사용하는 방법을 선택했다. 노이즈 제거 알고리즘은 침식과 팽창을 이용한 열기, 닫기를 사용했다.

2.4 정상 객체 마스킹 기법

본 논문에서 제안하는 분할 네트워크를 이용한 그림 8이 설명하고 있는 정상객체 마스킹 단계를 아래에 자세히 설명한다. 카메라 영상에서 불규칙적으로 이동하는 정상 객체가 포함된 입력 이미지의 경우 제안하는 객체 마스킹 기법을 사용시 오토인코더가 인코딩 된 벡터를 복구하여 만든 출력에서 목표 정상 객체를 마스킹하여 제거한다. 이와 같이 제거된 정상 객체는 잔차 오류 이미지에 나타나지 않으며, 비정상 객체와 이상동작 객체만 오류 이미지에 나타나게 된다. 입력 영상에 대한 이상 여부의 최종 판단은 제곱 손실인 오류 이미지의 제곱의 합을 이용해 판단한다.

$$L2\_loss=\Sigma\{(Input - Reconstructed)^2 * Mask\} \quad (4)$$

$$Residual\_error\_image=(Input-Reconstructed)^2 * Mask \quad (5)$$

수식(4)와 수식(5)를 통해 비정상으로 판단한 정상 객체를 마스킹 함으로서 거짓 참을 감소시켜 정확도를 증가시킨다.

2.5 전체 네트워크 훈련

의사 라벨을 획득하고 이를 사용해 네트워크를 훈련하기 위해 모든 가중치를 한 번에 훈련시키지 않고 아래

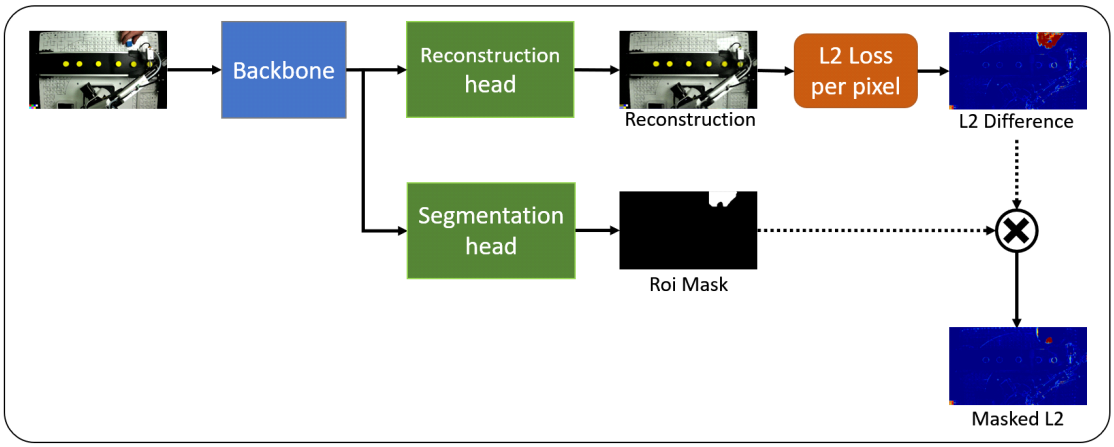


그림 8. 특정 객체 마스킹
Fig. 8. Specified object masking.

와 같이 일련의 과정을 거쳐 부분별로 훈련시킨다.

2.5.1 오토인코더 훈련

첫번째 단계에서 분할 헤드를 훈련시킬 데이터가 존재하지 않기 때문에 이를 획득하기 위해 백본과 디코더 헤드를 먼저 훈련시킨다. 백본이 빨강, 초록, 파랑 3가지 색상에 대한 데이터를 부족하게 학습할 시 해당 색상이 비정상 입력으로 들어왔을 시 정상인 부분을 복구하지 못할 가능성이 증가한다[32]. 이 문제를 개선하기 위해 다양한 이미지로 사전 학습된 가중치[33]를 사용하거나, 학습 초기에 color jitter[34]증강을 사용한다.

2.5.2 분할 라벨 생성 및 훈련

상기 방법으로 훈련된 백본과 디코더 헤드를 이용하여 2.2에서 제안한 방법으로 정상 객체에 대한 마스크 라벨을 획득한다. 획득한 라벨을 정답으로 사용해 분할 헤드를 훈련시킨다.

2.5.3 오토인코더 교정

2.5.2의 과정에서 백본을 동결할 수 있을 만큼 충분한 성능이 달성된 경우에는 아래 과정은 생략할 수 있다. 그러나 대부분의 경우에는 2.4.2의 과정에서 백본의 가중치가 2.4.1을 실행했을 때와 차이가 생겨 디코더 헤드의 결과 이미지 복구 성능이 크게 감소한다. 이를 보정하기 위해 백본과 분할 헤드를 동결하고 디코더 헤드만 적은 반복으로 훈련시켜서 오토인코더의 정확도를 보정한다.

Ⅲ. 실 험

3.1 테스트 환경

앞에서 제안한 통합 합성곱 신경망을 테스트하기 위해 로봇 팔과 컨베이어 벨트로 만들어진 환경을 구현했다. 배경만을 포함한 정상 데이터는 컨베이어 벨트와 로봇 팔이 노란색 원통을 지속적으로 순환시킨다. 정상 객체가 포함된 정상상황은 정상 객체인 손을 움직이는 것으로 한다. 컨베이어 벨트 외부에 손을 제외한 객체가 있을 시 비정상으로 분류한다.

3.2 데이터 생성

라벨 자동생성 기법을 테스트하기 위해 정상객체와 비정상 객체를 이용해 데이터 생성을 했다. 정상 객체 및 비정상 패치를 정상 데이터에 삽입함으로써[35] 테스트를 위한 다양한 사례를 제작하였으며, 삽입한 패치의

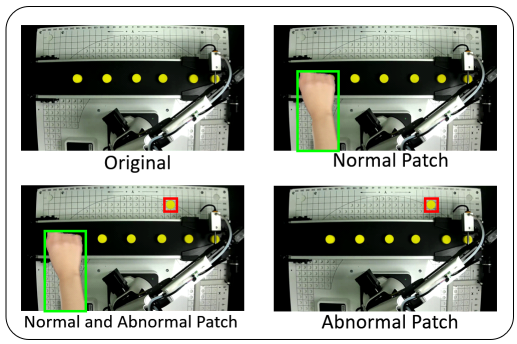


그림 9. 정상, 비정상 패치를 통한 데이터 생성
Fig. 9. Data augmentation through normal and abnormal patches

정확한 형태와 위치를 알고 있기 때문에 이를 평가하기 위한 정답으로 사용한다. 그림 9는 데이터 생성 결과이다. 테스트용 데이터를 위해 294장의 프레임을 가진 손을 촬영한 영상에 수작업으로 라벨링을 했다. 각 객체는 평균적으로 300k개의 픽셀을 가진다. 라벨링으로부터 얻은 객체를 그림 9의 방법으로 이상감지 훈련 데이터에 증강을 진행하여 정상, 비정상 객체가 포함된 훈련 데이터와 평가 데이터를 획득했다.

3.3 오토인코더를 통한 라벨 자동생성

3.2절에서 획득한 데이터를 오토인코더를 이용한 그림 7의 절차로 처리하여 의사 라벨을 획득한다. 표 .3은 정답과 자동 생성된 라벨을 사용한 제안된 모델의 훈련 결과의 객체분할 마스크 mAP 성능이다. 라벨 자동생성으로 획득한 의사 라벨[35]로 훈련한 가중치는 단독으로 사용하였을 때 테스트 데이터에서 정답 대비 마스크 mAP.50에서 96.82%의 성능을 보인다.

본 논문에서 제안한 라벨 자동생성은 Intel core-i7 4GHz, GPU Nvidia GTX3090 기반 하드웨어 환경에서 초당 1.24개의 처리 속도(장당 0.81초)를 가진다. 이는 마스크 폴리곤 라벨링보다 비교적 간단한 경계상자 라벨링 1장을 사람이 하는데 필요한 시간인 88.0초[34]보다 라벨링 속도를 11배 향상시키는 결과를 제공함을 알 수 있다.

3.4 이상감지

실험을 위해 그림 10의 1개의 로봇 팔과 컨베이어 벨트가 포함된 테스트 환경을 구성했다. 훈련용으로 정상 데이터 9283장과 정상 객체를 포함한 라벨이 없는 1057장을 사용하였으며, 평가용으로 총 1000장을 사용했다. 각 상황에 대한 데이터 수를 표 4에 나타낸다. 기존 오토인코더를 사용한 방법은 비정상 상황을 정상

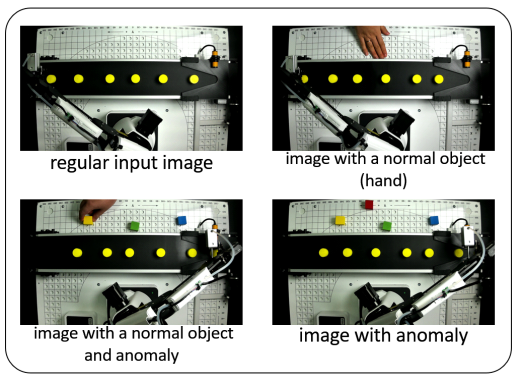


그림 10. 테스트 환경
Fig. 10. Test environment

표 4. 테스트용 데이터셋
Table 4. Test datasets

	All	Background only	Normal object	Anormal object	Normal and abnormal object
Train	10340	9283	1057		
Validation	1000	250	250	250	250

으로 판단하는 것을 예방하기 위해 학습된 영상과 유사한 영상만을 출력으로 복구하게 훈련을 한다. 이 때문에 훈련 시 정상영상으로 라벨링된 물체의 위치나 방향에 작은 변화라도 차이가 있는 테스트 영상의 경우 이를 복구하지 못하여 오동작 영상으로 판단한다.

기존의 오토인코더와 비교 대비 본 논문에서 제안한 모델이 정상객체가 포함된 정상 상황의 정확도를 크게 개선하는 결과를 보인다. 그러나 정상 객체가 포함된 정상상황의 이상감지 결과는 표 3의 마스크 감지 결과와 표 5의 배경만 존재하는 정상 상황 대비 낮은 정확도인 66.4%만을 달성할 수 있었다. 이는 정답과 추론한 마스크 범위 전제 대비 공통 부분이 임계 값 보다 큰 모든 경우를 정답으로 처리하는 객체 분할과 다르게 이상 감지가 추론한 마스크가 같거나 큰 경우를 정답으로 처리하기 때문으로 분석되었다. 이에 대한 예시로 그림 11의 손실지도 (Loss Map)의 마스크가 실제 정상 객체를 전부 포함하지 못하여 마스킹된 손실지도 (Loss Map)에 작업자의 손 (특정 객체)의 윤곽선이 남은 것을 볼 수 있다. 이를 개선하기 위해 2가지 방법을 적용한다. 방법1 마스크에 확장 적용, 방법.2 마스크 범위에 포함된 픽셀의 손실 값에 비례하여 전체 손실 값에 편향

표 5. 통합 CNN을 이용한 이상감지
Table 5. Anomaly detection using integrated CNN

	All	Background only	Normal object	Anormal object	Normal and abnormal object
Autoencoder	69.10	77.20	0.00	99.20	100
Ours	84.70	78.00	66.40	98.40	**95.20**
Ours (erosion mask 3)	83.90	78.00	70.00	98.40	89.20
Ours (erosion mask 5)	83.80	78.00	72.40	98.40	86.40
Ours (biased loss)	**85.00**	78.00	**91.60**	98.40	72.00

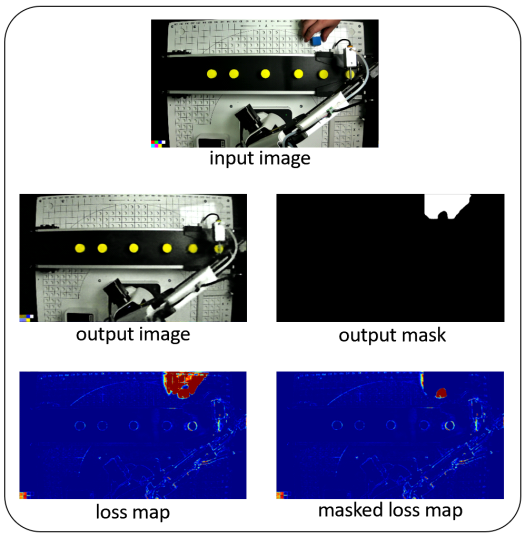


그림 11. 통합 CNN을 이용한 정상객체 마스킹
Fig. 11. Normal object masking using Multi-Task CNN

적용. 방법.2 사용시 특정 객체 마스킹을 통한 필터링 정확도가 91.60%로 상승하였다. 하지만 해당 과정을 거침으로서 마스크의 크기가 확대되어 비정상 상황에서의 정확도가 감소하는 문제가 발생했다.

객체 분할 그림 11은 테스트용 데이터셋의 헤드 별 추론 결과와 최종 네트워크 출력인 마스킹된 손실 지도이다.

Ⅳ. 결 론

본 논문에선 오토인코더와 분할 네트워크의 장점을 강화하고 단점을 보안하기 위해 두 네트워크를 통합한 특정 객체 필터링 및 이상동작 감지를 위한 통합 합성곱 신경망을 제안한다. 비지도 학습과 마스크 라벨 자동생성 기법을 사용해 대부분의 훈련 과정을 자동화한다. 이를 통해 라벨링에 걸리는 시간을 11배 단축시킨다. 또한 정상 객체 마스킹은 기존에 오토인코더의 성능에 상당한 개선을 제공한다. 제한되는 상황인 불규칙적 움직임을 가진 정상 객체를 마스킹하여 전체 정확도를 15.90% 증가시킨다. 이러한 실험 결과는 기존 비지도 학습 이상 감지가 가지는 한계를 극복해 더 나은 성능을 가질 수 있다는 것을 보여준다. 또한 이상 감지 뿐만 아니라 오토인토더의 라벨 자동생성의 가능성을 확인할 수 있다. 하지만 수작업으로 만든 정답 라벨에 비교해 정확도 손실이 있다는 제한사항이 존재한다. 이를 개선해 이상감지의 정확도를 증가시키시 위한 향후 연구를 제안한다.

References

[1] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” *arXiv preprint arXiv:2003.05991*, 2020. (https://doi.org/10.48550/arXiv.2003.05991)

[2] M. Sakurada and T. Yairi, “Anomaly detection using autoencoders with nonlinear dimensionality reduction,” in *Proc. MLSDA 2014 2nd Wkshp. Mach. Learn. Sensory Data Anal.*, pp. 4-11, 2014. (https://doi.org/10.1145/2689746.2689747)

[3] S.-K. Kang, M.-H. Park, Y.-H. Kim, N.-W. Kim, and I.-Y. Seo, “Development of anomaly-detection system for the underground cable tunnel using autoencoder,” *The Trans. The KIEE*, vol. 69, no. 2, pp. 69-75, 2020. (https://doi.org/10.5370/KIEEP.2020.69.2.69)

[4] E. Eskin, et al., “A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data,” *Applications of Data Mining in Comput. Secur.*, pp. 77-101, 2002. (https://doi.org/10.1007/978-1-4615-0953-0_4)

[5] W. Bailer and H. Fassold. “Resource-efficient object detection by sharing backbone CNNs,” *2019 IEEE ISM*, pp. 196-1963, 2019. (https://doi.org/10.1109/ISM46123.2019.00042)

[6] D.-H. Lee, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,” *Wkshp. Challenges in Representation Learn., ICML.*, vol. 3, no. 2, 2013. (https://www.kaggle.com/blobs/download/forum-message-attachment-files/746/pseudo_label_final.pdf)

[7] H. B. Barlow, “Unsupervised learning,” *Neural Computation*, vol. 1, no. 3, pp. 295-311, 1989. (https://doi.org/10.1162/neco.1989.1.3.295)

[8] A. T. Sufian, et al., “A roadmap towards the smart factory,” *2019 12th Int. Conf. Develop. Esystems Eng. (DeSE)*, IEEE, pp. 978-983, 2019. (https://doi.org/10.1109/DeSE.2019.00182)

[9] M. M. Mabkhot, et al., “Requirements of the

smart factory system: A survey and perspective," *Machines*, vol. 6, no. 2, 23, 2018.
(https://doi.org/10.3390/machines6020023)

[10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surv. (CSUR)*, vol. 41, no. 3, pp. 1-58, 2009.
(https://doi.org/10.1145/1541880.1541882)

[11] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, pp. 21-49, 2008.
(https://doi.org/10.1007/978-3-540-75171-7_2)

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84-90, 2017.
(https://doi.org/10.1145/3065386)

[13] B. Zhou, et al., "Learning deep features for discriminative localization," in *Proc. IEEE Conf. CVPR*, pp. 2921-2929, 2016.
(https://doi.org/10.48550/arXiv.1512.04150)

[14] A. R. Ajiboye, R. Abdullah-Arshah, and Q. Hongwu, "Evaluating the effect of dataset size on predictive model using supervised learning technique," *IJSECS*, vol. 1, pp. 75-84, Feb. 2015.
(http://dx.doi.org/10.15282/ijsecs.1.2015.6.0006)

[15] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1-54, 2019.
(https://doi.org/10.1186/s40537-019-0192-5)

[16] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
(https://doi.org/10.48550/arXiv.1312.6114)

[17] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special lecture on IE*, vol. 2, no. 1, pp. 1-18, 2015.
(http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf)

[18] D. Gong, et al., "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proc. IEEE/CVF Int. Conf. Computer Vision*, pp. 1705-1714, 2019.
(https://doi.org/10.48550/arXiv.1904.02639)

[19] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2017.
(https://doi.org/10.1145/3097983.3098052)

[20] D. T. Nguyen, et al., "Anomaly detection with multiple-hypotheses predictions," *Int. Conf. Mach. Learn., PMLR*, pp. 4800-4809, 2019.
(https://doi.org/10.48550/arXiv.1810.13292)

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th Int. Conf.,* Munich, Germany, Oct. 2015, Proceedings, Part III 18, Springer International Publishing, 2015.
(https://doi.org/10.48550/arXiv.1505.04597)

[22] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *Int. J. Multimedia Inf. Retrieval*, vol. 9, no. 3, pp. 171-189, 2020.
(https://doi.org/10.1007/s13735-020-00195-x)

[23] D. Bolya, et al., "Yolact: Real-time instance segmentation," in *Proc. IEEE/CVF Int.. Conf. Computer Vision*, pp. 9157-9166, 2019.
(https://doi.org/10.48550/arXiv.1904.02689)

[24] H. Liu, et al., "Yolactedge: Real-time instance segmentation on the edge," *2021 IEEE ICRA*, pp. 9579-9585, 2021.
(https://doi.org/10.1109/ICRA48506.2021.9561858)

[25] T.-Y. Lin, et al., "Feature pyramid networks for object detection," in *Proc. IEEE Conf. CVPR*, pp. 2117-2125, 2017.
(https://doi.org/10.48550/arXiv.1612.03144)

[26] K. He, et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. CVPR*, pp. 770-778, 2016.
(https://doi.org/10.48550/arXiv.1512.03385)

[27] T. N. Sainath, B. Kingsbury, and B.

Ramabhadran, “Auto-encoder bottleneck features using deep belief networks,” *2012 IEEE ICASSP*, pp. 4153-4156, 2012. (https://doi.org/10.1109/ICASSP.2012.6288833)

[28] H. Zhao, et al., “Loss functions for image restoration with neural networks,” *IEEE Trans. Computational Imaging*, vol. 3, no. 1, pp. 47-57, 2016. (https://doi.org/10.1109/TCI.2016.2644865)

[29] W. Burger, et al., *Principles of digital image processing,* vol. 111, London: Springer, 2009. (https://doi.org/10.1007/978-1-84800-195-4_2)

[30] K. R. Castleman, *Digital image processing,* Prentice Hall Press, 1996. (https://doi.org/10.1007/978-3-662-03477-4_4)

[31] T. H. Reiss, ed., *Recognizing planar objects using invariant image features,* Berlin, Heidelberg: Springer Berlin Heidelberg, 1993. (https://doi.org/10.1007/BFb0017560)

[32] Y. Li and N. Vasconcelos, “Repair: Removing representation bias by dataset resampling,” in *Proc. IEEE/CVF Conf. CVPR*, pp. 9572-9581, 2019. (https://doi.org/10.48550/arXiv.1904.07911)

[33] X. Ying, “An overview of overfitting and its solutions,” *J. Physics: Conf. series*, vol. 1168, p. 022022, IOP Publishing, 2019. (https://doi.org/10.1109/cvpr.2019.00980)

[34] B. Neyshabur, et al., “Exploring generalization in deep learning,” *Advances in NIPS*, vol. 30, 2017. (https://doi.org/10.48550/arXiv.1706.08947)

[35] G. Ghiasi, et al., “Simple copy-paste is a strong data augmentation method for instance segmentation,” in *Proc. IEEE/CVF Conf. CVPR*, pp. 2918-2928, 2021. (https://doi.org/10.1109/cvpr46437.2021.00294)

[36] H. Su, J. Deng, and L. Fei-Fei, “Crowdsourcing annotations for visual object detection,” *Wkshp. Twenty-Sixth AAAI Conf. Artificial Intell.*, 2012. (http://ai.stanford.edu/~haosu/papers/bbox_submission.pdf)

홍 상 욱 (Sang-wook Hong)

2021년 2월 : 충북대학교 전자공학부 학사 졸업
2023년 2월 : 충북대학교 전자공학부 석사 졸업
<관심분야> 전자공학, Deep learning, Image Recognition, Self-Supervised Learning

김 형 원 (Hyung-won Kim)

1991년 2월 : KAIST 전기 및 전자공학과 학사 졸업
1993년 2월 : KAIST 전기 및 전자공학과 석사 졸업
1999년 8월 : University of Michigan, Ann Arbor, Electrical and Computer Engineering, Ph.D.
1999년 8월 : Synopsys, Mountain View, CA, USA
2001년 1월 : Broadcom, San Jose, CA, USA
2005년 10월 : (주)카이로넷, Founder & CEO
2013년 2월~현재 : 충북대학교, 전자공학부, 교수
<관심분야> Deep learning, Image Recognition, Self-Supervised Learning, AI Accelerator Architecture and Chip Design, Low Power Circuits, Wireless Communications
[ORCID:0000-0003-2602-2075]