Journal of KIIT. Vol. 22, No. 7, pp. 1-11, Jul. 31, 2024. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2024.22.7.1

효과적인 족적 감정 수사를 위한 자기 지도 학습 및 이진화 기반 이미지 검색 기술

이창엽*[1], 김동주*[2], 서영주*[3], 황도경*[4]

Self-Supervised Learning and Binarization-based Image Retrieval Technology for Effective Forensic Footprint Analysis

Chang-Yeop Lee*[1], Dong-Ju Kim*[2], Young-Joo Suh*[3], and Do-Kyung Hwang*[4]

본 논문은 2024년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업이며 (No. 2022R1A6A1A0305295413), 과학기술정보통신부·경찰청이 공동 지원한 '폴리스랩 2.0 사업'의 (RS-2023-00281072) 지원을 받아 수행된 연구임

요 약

본 논문에서는 족적 이미지 검색 분야에서 라벨링 되지 않은 원시 데이터를 이용해 효율적으로 특징을 추출하기 위한 새로운 학습 방법을 제안한다. 이 분야는 현장에서 촬영한 이미지와 사전 정의된 이미지를 각각의 딥러닝 모델에 통과시켜, 그 결과물들을 비교하는 방식이 사용된다. 따라서, 유사한 결과물을 추출하기 위해 네트워크의 학습 방법이 중요하며, 기존 연구는 주로 샴 네트워크를 사용하고 있었다. 그러나, 샴 네트워크는 수많은 데이터에 대한 라벨링이 필요하고 이에 따른 인적 자원과 비용이 든다는 제한점이 존재했다. 이를 극복하기 위해, 본 논문은 원시 데이터만으로 효율적으로 학습이 가능한 자기 지도 학습을 제안하였다. 이에 대한 최적화를 가속하기 위한 이진화 작업을 도입하였으며, 제안하는 방법이 이미지 검색의 전통적인 방식인 샴 네트워크보다 우수한 성능을 기록하는 것을 실험으로 입증하였다.

Abstract

This paper proposes a novel learning method for efficiently extracting features from raw, unlabeled data in the field of footprint image retrieval. In this field, images captured on-site and predefined images are passed through deep learning models, and their outputs are compared. Therefore, the learning method of the network is crucial for extracting similar results. Previous studies have primarily used Siamese networks. However, Siamese networks require extensive data labeling, which incurs significant human resources and costs. To overcome this limitation, this paper suggests a self-supervised learning approach that can efficiently learn using only raw data and introduces a binarization process to accelerate optimization. Experiments demonstrated that the method proposed in this paper outperforms the traditional Siamese network approach in image retrieval.

Keywords

image retrieval, deep learning, computer vision, self-supervised learning, footprint

* 포항공과대학교 인공지능연구원(*4 교신저자)
- ORCID[1]: https://orcid.org/0009-0003-0607-9901
- ORCID[2]: https://orcid.org/0009-0009-6950-4200
- ORCID[3]: https://orcid.org/0000-0001-7208-1709
- ORCID[4]: https://orcid.org/0000-0003-4271-5672

· Received: Apr. 16, 2024, Revised: Jun. 20, 2024, Accepted: Jun. 23, 2024
· Corresponding Author: Do-Kyung Hwnag
Dept. of POSTECH Institute of Artificial Intelligence,
Cheongam-ro 77, Nam-gu, Pohang, Korea
Tel.: +82-54-279-5663, Email: dokyung@postech.ac.kr

Ⅰ. 서 론

범죄 현장에서 족적은 물리적 증거 중 하나로, 범죄 사건 조사에서 중요한 단서의 역할을 한다[1]-[3]. 기존에는 수사관들이 범죄 현장에서 족적 이미지를 수집하고 데이터베이스에서 가장 유사한 신발 이미지를 수동으로 비교하며 단서를 찾았다. 하지만 족적 이미지와 데이터베이스의 신발 이미지의 양이 기하급수적으로 증가함에 따라 수동으로 비교하기는 많은 제한점이 있고, 자동으로 신발에 대한 정보를 검색하는 방법에 대한 것이 연구 과제로 남아있었다.

이러한 침체 속에서 최근 몇 년 동안, 딥러닝 기술 발전은 많은 연구 분야에서 혁신을 가져왔다[4]-[6]. 이에 따라, 족적 이미지를 자동으로 검색하는 것과 같이 복잡한 문제를 해결하는 데도 딥러닝이 중요한 역할에 기여하고 있고 많은 연구가 이루고 있었다[7]-[11]. 딥러닝을 이용한 족적 이미지에 대한 신발 정보를 자동 검색하는 일반적인 방법은 그림 1과 같다.

그림 1에서 쿼리 이미지는 범죄 현장에서 수집한 족적 이미지이고, 참조 이미지 데이터베이스는 쿼리 이미지와 비교할 수많은 신발 정보를 모은 데이터들의 집합이다. 먼저, 쿼리 이미지와 참조 이미지들이 각각 딥러닝의 신경망으로 구성된 특징 추출 모델로 입력되고, 이에 대한 출력은 각각의 특징을 잘 나타내는 벡터 형태로 반환된다. 그 후, 쿼리 이미지의 특징 벡터가 참조 이미지들의 특징 벡터들과 Euclidean 거리로 유사도를 계산한다. 최종적으로, 유사도가 높은 순서로 랭킹을 매겨 쿼리 이미지와 참조 이미지 간에 최적의 쌍을 찾는 절차이다.

이러한 절차는 단순하게 딥러닝 네트워크에 이미지를 통과시켜 각 이미지가 표현하고 있는 특징으로 카테고리를 출력하는 이미지 분류 분야를 넘어, 쿼리 이미지와 참조 이미지를 각각 딥러닝 네트워크를 통과시켜 출력된 특징 간의 유사도를 높이는 이미지 검색 분야로 구분된다. 이 때문에 특징 추출 모델의 학습 방법이 중요해졌으며, 대표적으로 샴 네트워크가 사용된다[12]. 이 네트워크는 쿼리 이미지를 기준으로 관련이 있는 이미지는 양성 이미지, 관련이 없는 이미지는 음성 이미지로 별도의 라벨링을 한 후, 각각의 딥러닝 네트워크로 입력시키고 이미지를 검색하는 학습 방법으로 우수한 성과를 거두었다[13][14].

하지만, 범죄 현장에서 수집한 방대한 양의 족적 이미지들의 경우 대부분이 라벨링 되어 있지 않은 원시 데이터이며, 모든 데이터를 샴 네트워크와 같이 라벨링 하기엔 수많은 인적 자원과 비용이 든다.

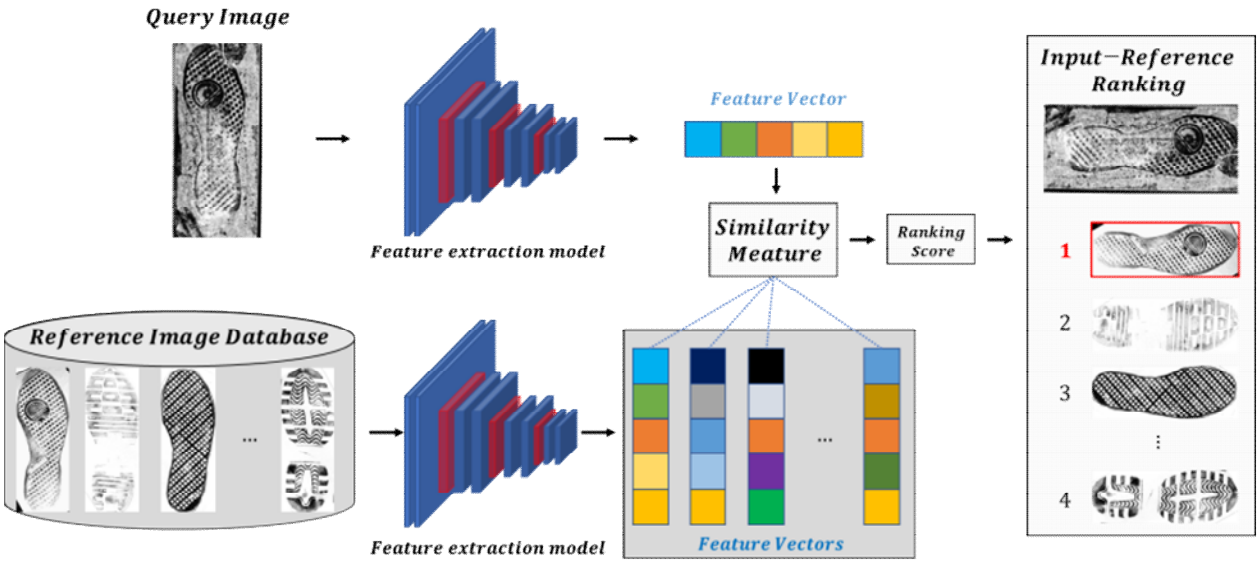

그림 1. 족적 이미지를 자동 검색하는 절차
Fig. 1. Procedure for automatic footprint image retrieval

이러한 제한 사항 중, 최근 딥러닝 분야에서 라벨링 데이터 없이 원시 데이터만으로 강력하고 견고한 특징을 추출하여 최첨단의 성능을 가지는 자기 지도 학습 대한 많은 연구가 이루어지고 있고, 그 우수성이 입증되고 있다[15]-[19].

이러한 이유로, 본 논문에서는 기존과 같이 샴 네트워크를 적용하는 것이 아닌, 족적에 대한 수많은 원시 데이터로 자기 지도 학습을 하여 효율적으로 특징을 추출하는 학습 방법을 제안한다. 아울러 족적이 포함되어 있지 않은 불필요한 부분은 제거하고 하고, 패턴을 이진화하는 과정을 선행적으로 진행하여 추가적으로 학습에 대한 최적화를 가속한다. 이러한 학습 방법으로 특징을 효율적으로 추출해, 기존의 족적 이미지 검색보다 우수한 성능을 기록했음을 경찰청 데이터 및 공인 데이터 세트에서 실험으로 입증하였다.

II. 관련 연구

2.1 기존 족적 데이터에 대한 이미지 검색 학습 네트워크

기존 족적 이미지 검색에서 학습 네트워크로는 샴 네트워크가 지배적이였다. 샴 네트워크는 입력 데이터, 입력데이터와 같은 클래스인 양성 데이터, 다른 클래스인 음성 데이터로 총 3개의 스트림으로 구성된다. 이 각각의 스트림은 모두 동일한 신경망 구조와 가중치를 공유하고 최종적으로 Triplet Loss를 계산하는 방식으로 학습이 이루어진다[12]-[14]. 기존 연구들은 이 샴 네트워크의 학습 방법을 기반으로 최적의 모델을 설계하기 위해 많은 관심을 기울이고 있었다.

Z. Ma et al.은 두 개의 스트림 구조를 활용하여, 해당 하위 이미지가 포함하는 의미 있는 정보의 양을 반영하여 가중치를 부여하는 MP-CNN(Multi-Part weighted Convolutional Neural Network)를 제안하여 수렴을 가속화시켰다[7]. W. Liu and D. Xu는 딥 해시 네트워크로 족적 이미지를 이진화시키고, 동시에 STN(Spatial Transformer Network)를 내장시켜 회전된 족적 이미지에도 정확도를 개선시키는 네트워크를 개발하였다[8]. J. Dai et al.[20]은 연산속도와 메모리 사용량에 이점이 있는 Deformable Convolution을 기반하여 두 가지 모듈을 구성하였다. SA(Spatial Attention) 모듈에서 저수준의 특징과 CA(Channel Attnetion) 모듈에서 고수준 특징을 추출하고, 두 모듈을 결합한 DALH-CNN(Dual Attention Light Hash CNN) Network를 새롭게 제안하며 정확도와 속도를 크게 향상시켰다[9].

위와 같이, 샴 네트워크를 기반으로 한 많은 연구들에 대한 우수성이 입증되었다. 하지만, 이 방법은 족적에 대한 방대한 양의 원시 데이터에 대한 양성 데이터 및 음성 데이터에 대한 라벨링 작업이 필요하며, 이는 수많은 인적 자원과 비용이 든다는 제한점이 있다.

2.2 자기 지도 학습

최근 라벨링 되지 않은 원시 데이터들로 특징을 효과적으로 추출하기 위해 자기 지도 학습에 대한 연구가 많이 이루어지고 있다.

W. Su et al.는 다양한 모달리티/출처의 데이터와 자기 지도 학습의 사전 학습 방법론을 적용하는 M3I Pre-trainig(Maximizing Multi-modal Mutual Information Pre-training)를 제안하며 이미지 분류, 객체 탐지, 장기 객체 탐지 등 다양한 도메인에서 우수한 성능을 내었다[15]. P. Wang et al.은 시각, 오디오, 언어 모달리티 간 표현을 원활하게 정렬하는 모달리티 어댑터를 제안하고, 공유 자기 주의 계층을 구성하여 다중 모달 융합 하여 단일 및 다중 모달 작업에서 선도적인 결과를 달성하였다[16]. S. Srivastava and G. Sharma는 다중 모달리티를 통합하기 위해 자기 지도 마스크 훈련 방식으로 사전 학습 뒤, 작업 특정 인코더, 모든 작업에 공통된 트렁크, 작업 특정 예측 헤드를 구성함으로써 다양한 작업에 일반화 성능을 보여주었다[17]. K. He et al.은 새로운 자기 지도 학습 방식인 MAE(Masked Auto Encoder)를 제안하며 많은 사전 학습 연구에 기여해왔다[18]. 이 구조는 비대칭적인 인코더 및 디코더의 네트워크와 이미지 패치를 70% 이상 마스크하며 사전 학습을 진행했으며, 훈련 속도를 3배 이상 가속화하고, 기존 지도기반 사전 학습을 능가하는 결과를 가져왔다.

본 논문은 위와 같은 방법론 중, 단일 모달리티인 이미지에 대해 자기 지도 학습 방식을 적용한 MAE를 채택하고, 족적에 대한 수많은 원시 데이터에서 특징을 효과적으로 추출하기 위해 활용한다.

III. 특징 추출을 위한 학습 방법

본 논문에서 제안하는 학습 방법은 특징을 추출하기 위한 자기 지도 학습과 이를 가속화하기 위한 이진화에 대한 방법으로 나뉘며, 학습 및 평가를 위한 데이터는 대한민국 경찰청에서 제공받은 현장 족적 이미지 그리고 등록 이미지를 각각 활용하였다.

3.1 MAE(Masked Auto-Encoder)

현장에서 수집한 족적에 대한 원시 데이터만으로 자기 지도 학습을 하기 위해, MAE의 학습 방식을 활용한다[18]. 이 방식은 이미지에 대한 단일 모달리티만으로 자기 지도 학습을 가능하게 한 모델이며, 그림 2은 원시 데이터에 MAE를 적용한 전체 학습 과정을 나타낸 것이다.

우선, 족적 이미지를 패치단위로 분할한다. 이를 위해 컨볼루션 레이어의 커널 크기를 패치 크기와 맞추고, 패치 간격에 맞게 스트라이드를 설정한다. 이렇게 분할된 이미지 패치들을 높은 비율(e.g. 75%)로 랜덤하게 마스킹하고, 마스킹되지 않은 나머지 이미지 패치들만이 인코더로 입력된다. 인코더는 ViT(Vision Transformer) 구조로, Self-Attention 연산을 통해 각 패치간의 유사도를 비교하여 특징을 학습하고, FFN(Feed-Forward Network)를 사용해 학습된 특징들을 확장하며 다양한 표현을 학습해 나간다[21]. 이러한 구조는 전역적 수용 필드를 가지므로, 지역적 수용 필드를 가지는 기존 모델들보다 더 풍부한 특징을 학습할 수 있다[22][23]. 풍부한 특징을 추출한 이미지 패치들에 마스킹된 패치들을 다시 결합하고, 이 결합된 패치들이 디코더로 입력된다. 디코더는 원본 이미지를 재구성하는 목적으로만 설계되었고, 학습이 끝난 후 이미지 검색에서 족적에 대한 특징 추출 모델로 사용할 때는 제거하기 때문에, 소수의 레이어들로만 구성한다.

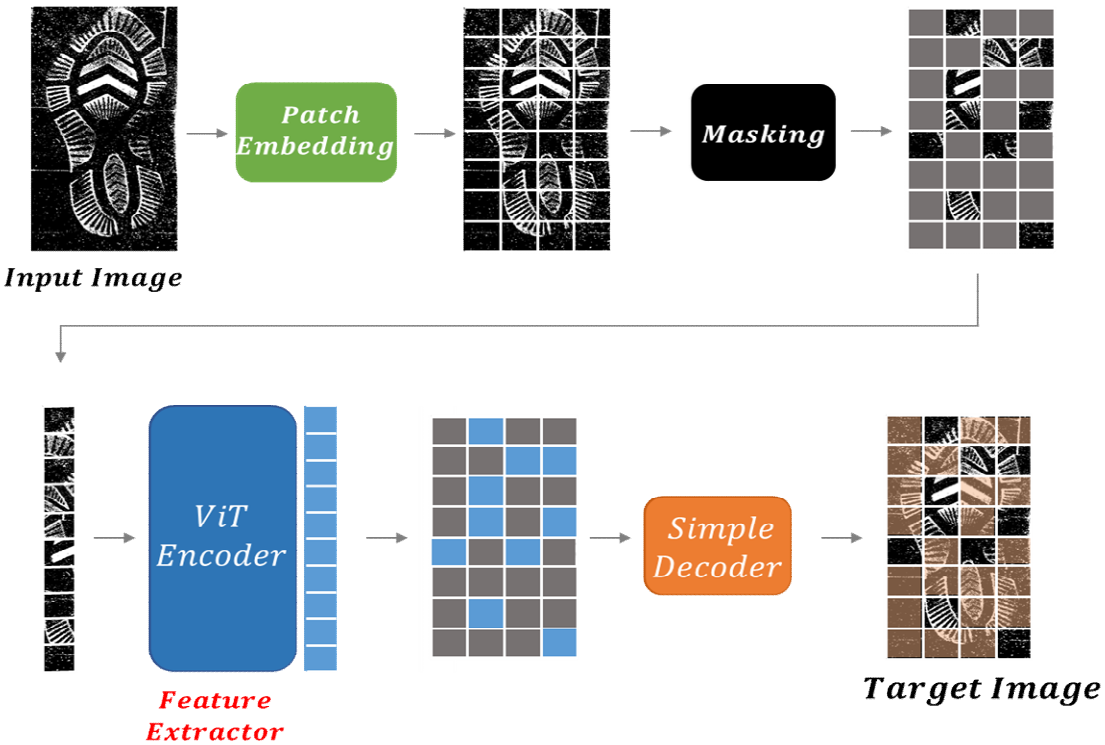

그림 2. 마스크 된 오토 인코더로 족적 이미지 사전 학습
Fig. 2. Pre-training footprint images with masked autoencoder

이 과정에서 아주 낮은 비율(e.g 25%)의 이미지 패치들이 인코더에서 추출한 특징들만으로 디코더로 입력돼 원본 이미지를 재구성해야 하므로, 더 풍부한 특징을 학습할 수 있다.

위와 같은 방식으로 라벨링 되지 않은 원시 데이터를 이용한 방법이 라벨링 데이터를 이용한 기존 지도 기반의 학습 방식보다 더 우수한 성능을 내는 것이 증명되었다[18]. 본 논문은 이 방식을 활용해 경찰청에서 제공받은 방대한 양의 원시 데이터를 활용한 효율적인 특징 학습을 가능하게 하였고, 현장 족적 이미지와 사전 등록된 참조 이미지 간의 특징을 효과적으로 추출해 두 이미지 간의 유사도 차이를 기존의 방법보다 더 정확하게 하였다. 또한, 이러한 학습 방법은 라벨링 되지 않은 원시 데이터만을 활용하기 때문에, 수많은 인적 자원과 비용 또한 절감에 기여한다는 점에서도 유용하다.

3.2 이미지 이진화

추가적으로 학습에서 더 나은 최적화를 위해, 본 논문은 먼저 현장에서 수집된 쿼리 이미지와 및 사전 수집된 참조 이미지에 대해 3차원의 컬러 이미지를 1차원의 흑백 이미지로 이진화하는 과정을 거친다. 컬러 이미지는 흑백 이미지보다 높은 차원을 가져 계산 복잡도와 메모리 사용량에 부정적인 영향을 미칠 뿐만 아니라, 손실 함수에 대한 최적화에 어려움을 겪을 수 있다. 따라서, 현장의 쿼리 이미지 및 참조 이미지 모두 이진화가 적용된다.

이진화의 전체 과정은 그림 3과 같다. 먼저, 경찰청에서 수집한 현장 쿼리 이미지 및 사전 등록된 참조 이미지는 측정 도구와 함께 족적 이미지가 포함되어 있다. 이러한 불필요한 정보는 이진화에 부정적인 영향을 미칠 수 있으므로, 크롭을 통해 제거한다. 쿼리 이미지는 바닥에 명확하게 찍힌 자국이므로, OpenCV 모듈을 활용하여 간단하게 이진화를 진행하였다[24]. 그러나 참조 이미지의 경우, 자국이 선명한 현장 족적 이미지와 달리 실제 신발을 촬영한 이미지이기 때문에 입체감이 존재한다. 그러한 이유로 OpenCV 모듈을 활용하면 그림 3과 같이 선명하게 이진화가 되지 않는다는 제한점이 존재하였다. 이를 극복하기 위해, 본 논문에서는 신발의 깊이를 예측하고, 이를 기반으로 하는 Shoerinsics 방식을 사용해 입체감이 있는 신발을 선명하게 이진화를 가능하게 하였다[25].

이러한 이진화 과정을 통해 MAE 학습의 최적화가 가속화되었음을 실험 결과로 입증한다.

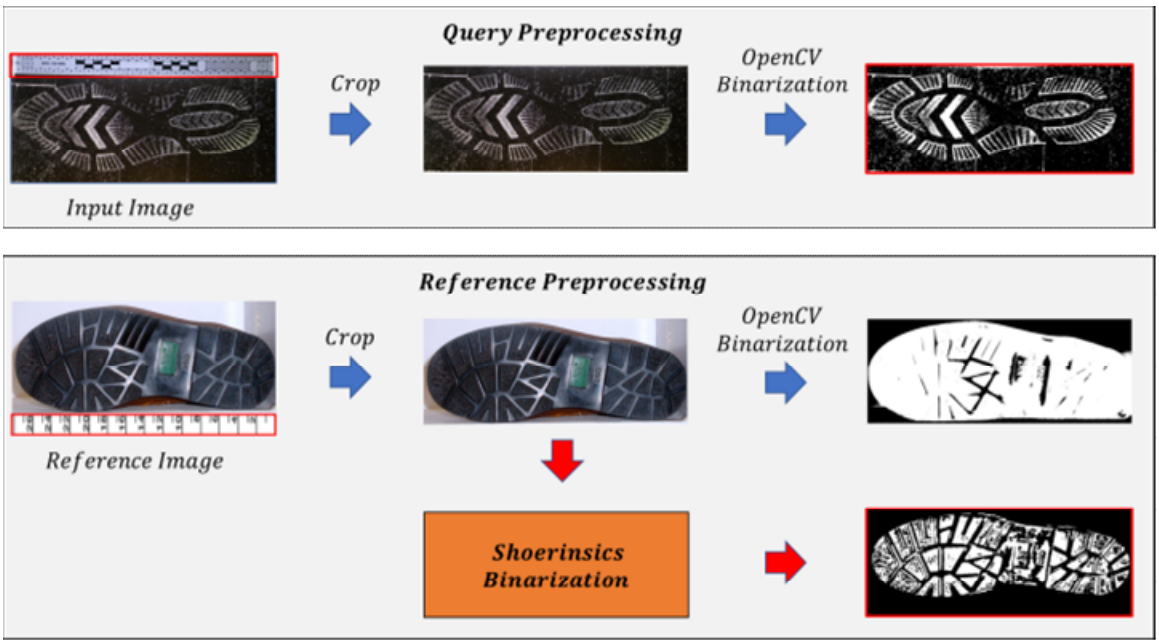

그림 3. 쿼리와 참조 이미지를 이진화하는 과정
Fig. 3. Process of binarization query and reference images

Ⅳ. 실 험

4.1 실험 환경 정의

본 논문에서는 특징 추출 모델을 학습하기 위해 경찰청에서 제공한 라벨링되지 않은 원시 데이터 약 20만개를 사용한다. 검증 데이터로는 공인 데이터 세트 FID-300과 경찰청 데이터를 사용하였으며, 각각 300개와 3772개로 구성되어 있다. 두 데이터 세트 모두 현장 족적 이미지와 사전 등록된 참조 이미지의 짝으로 이루어져 있다. 여기서 참조 이미지란, 현장에서 수집된 족적이 어떤 족적인지 판단할 수 있는 참고용 정답 데이터로 활용되며, 학습된 특정 추출 모델에 대한 성능 검증은 기존에 진행되었던 족적 이미지 검색 연구들과 동일하게 Top 1%, Top 5%, Top 10%를 평가 지표로 활용하였다[7]-[11]. 여기서 Top n%의 n은 허용 범위를 의미하는 숫자로써, 현장에서 수집된 이미지를 입력으로 할 때, 해당 입력과 유사하다고 판단되는 등록 이미지들의 범위의 제한을 의미한다. 즉, 전체 등록 이미지 개수가 1만장이라 가정을 할 때, Top 1% 기준, 쿼리 이미지와 유사하다고 판단되는 등록 이미지의 범위는 총 100장으로 구성되고, Top 10% 기준, 1,000장까지 허용된다. 실제 딥러닝 기반 족적 검색을 할 때, 입력과 유사하다고 판단되는 등록 이미지를 찾고, 모델이 판단한 등록 이미지 범위 중, 현장 족적과 일치한 등록 이미지가 해당 범위 안에 포함되어 있을 경우 식 (1)과 같이 True Positive로 검색 정확도를 산출한다. 식 1에서 i는 현장 이미지의 인덱스이고, y는 등록 이미지, X는 Top n%의 범위 안에서 등록 이미지와 유사하다고 판단된 이미지들의 집합이다. 아울러 모델의 견고함을 확인하기 위해, 모든 이미지들은 랜덤하게 증강되었고, 학습은 4 Tesla P40 환경에서 8 Batch size로 1000 Epochs만큼 학습되었다.

$$v_i = \begin{cases} 1 \text{ if } (y \in X) \\ 0 \; otherwise \end{cases} \quad (1)$$

4.2 학습된 피처 추출 모델의 성능 검증

먼저 정량적인 지표 확인으로, 표 1은 경찰청에서 수집한 데이터를 MAE 방식으로 학습된 특징 추출모델에 FID-300 데이터 세트의 현장 족적 이미지와 참조 족적 이미지를 각각 입력하여 유사도 비교 결과이다. Top 1%에서 73.8%, Top 5%에서 83.1%, Top 10%에서는 91.28%로 기존 모델 대비 모든 지표에서 가장 높은 성능을 기록하였다.

아울러 표 2는 위와 동일한 특징 추출 모델에 경찰청 제공 데이터 세트의 현장 족적 이미지와 참조 족적 이미지에 대한 유사도 비교 결과이고, Top 1%에서 60.3%, Top 5%에서 68.11%, Top 10%에서는 74.5%로 모든 지표에서 기존 모델들보다 높은 성능을 기록하였다.

표 1. FID-300 데이터 세트에 대한 성능 평가
Table 1. Performance evaluation on FID-300 dataset

Methods	Top 1%	Top 5%	Top 10%
Z. Ma et al.[7]	61.02	81.36	89.83
D. Li et al.[9]	66.67	83.54	89.92
Y. Wu et al.[10]	71.8	81.7	87.3
Ours	**73.8**	**83.1**	**91.28**

표 2. 경찰청 제공 데이터 세트에 대한 성능 평가
Table 2. Performance evaluation on dataset provided by the national police agency

Methods	Top 1%	Top 5%	Top 10%
Z. Ma et al.[7]	32.12	51.48	59.31
D. Li et al.[9]	31.97	54.14	58.22
Y. Wu et al.[10]	49.80	61.80	63.13
Ours	**60.30**	**68.11**	**74.50**

위와 같이 3개의 평가 지표에서 정량적인 결과뿐만 아니라 그림 4, 그림 5와 같이 시각적인 결과 또한 확인할 수 있다. FID-300와 경찰청 제공 데이터 세트에 대한 Top 1%에 해당하는 이미지는 각각 11개와 37개이고, Top 5%와 Top 10%는 더 많은 이미지를 가진다.

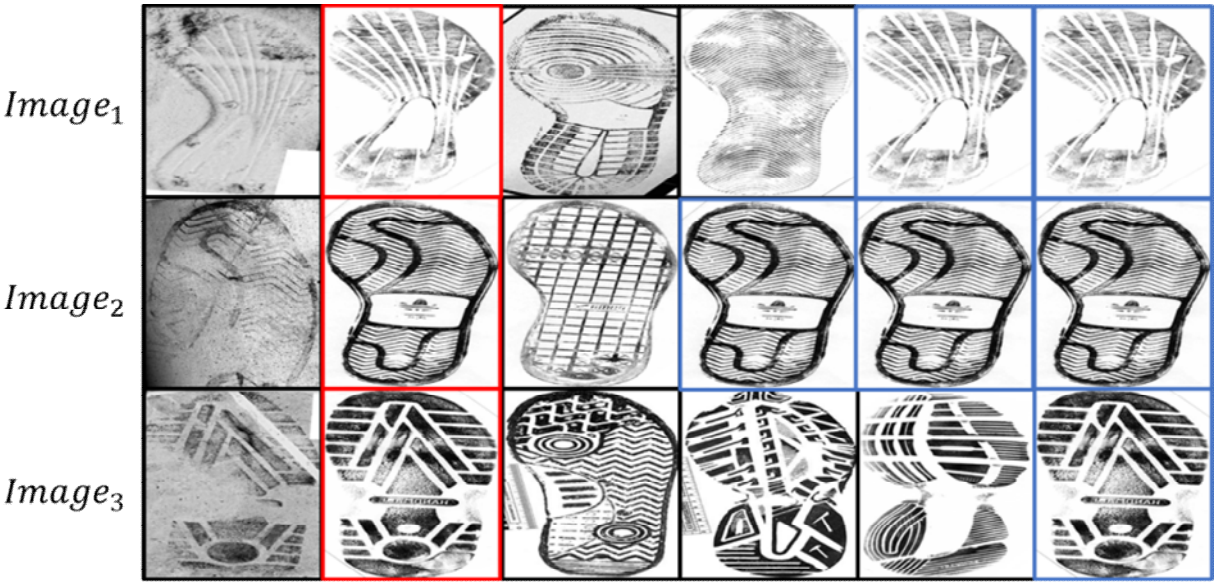

그림 4. FID-300 데이터 세트에서 기존 모델과 시각적인 비교 실험
Fig. 4. Visual comparative experiments with existing models on the FID-300 dataset

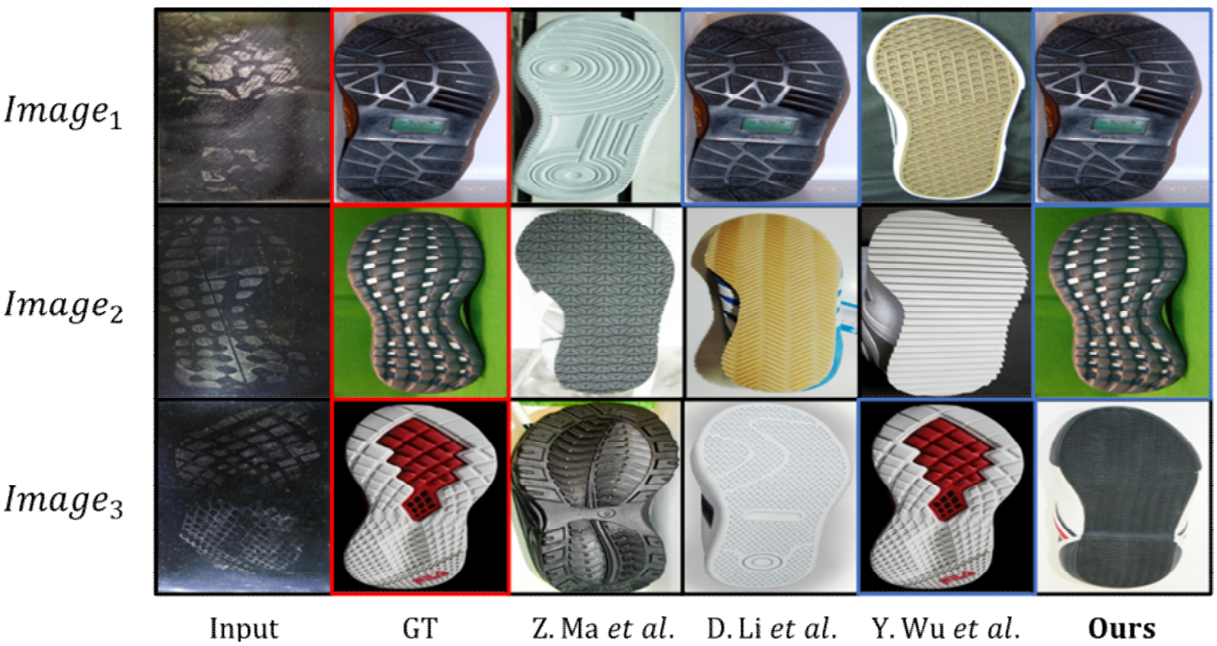

그림 5. 경찰청 제공 데이터 세트에서 기존 모델과 시각적인 비교 실험
Fig. 5. Visual comparative experiments with existing models on dataset provided by the national police agency

따라서, 모든 평가 지표에 대해 시각적인 결과를 제시하는 것은 어려움이 있으며, 각 모델의 Top 1%에서 11개, 37개 결과 중 3개만 시각화하였다. 그림 4 및 그림 5에서 첫 번째 열은 특징 추출 모델에 입력되는 현장 족적 이미지(Input)이고, 두 번째 열은 등록 이미지 (GT:Ground-Truth)이며, 나머지 열들은 각각의 모델이 출력하는 결과들이다.

먼저, FID-300 데이터 세트의 시각적 결과인 그림 4를 보면, 3개의 현장 이미지에 대해 Y. Wu et al.는 2개의 결과가 GT와 잘 매칭되었고, D. Li et al.는 1개, Z. Ma et al.는 하나도 매칭되지 못하였다. 반면, 본 논문에서 제안된 모델은 3개의 모든

결과가 잘 대응된 것을 볼 수 있다.

다음으로, 경찰청 제공 데이터 세트의 결과인 그림 5를 보면, Y. Wu et al.는 1개의 결과가 GT와 잘 매칭되었고, D. Li et al.는 1개, Z. Ma et al.는 하나도 매칭되지 못하였다. 반면, 본 논문에서 제안된 모델은 2개의 결과가 잘 대응된 것을 볼 수 있다.

이러한 결과로 미루어 보아, 제안된 모델은 기존의 모델들보다 현장 이미지와 참조 이미지 간의 특징 정보가 효과적으로 추출되었기에, 더 높은 정확성의 검색 결과가 나왔음을 확인할 수 있다. 하지만 그림 6과 같이 이미지 검색에 실패한 사례도 존재한다. 그림 4, 5는 바닥에 찍힌 족적 자국이 선명하게 남아있지만, 현장 환경에 따라 자국이 선명하지 않을 수 있다. 이에 따른 전처리나 네트워크 설계는 여전히 도전적인 과제로 남아있다.

그림 6. 이미지 검색 실패 사례
Fig. 6. Image search failed case

4.3 이미지 이진화

본 논문에서는 추가적으로 이미지의 이진화가 특징 추출에 미치는 영향에 대해 실험을 수행하였다. 그림 7은 이진화된 이미지와 이진화 처리가 되지 않은 컬러 이미지를 각각 전처리한 후, 1000 Epochs 동안 MAE를 사용하여 학습한 Loss 값의 변화를 시각화한 것이다. 결과적으로, 컬러 이미지의 Loss 값은 이진화된 이미지보다 0.077만큼 높게 나타나, 컬러 이미지가 최적화 과정에서 더 많은 어려움을 겪는 것으로 해석할 수 있다. 이는 컬러 이미지가 특징을 효과적으로 추출하지 못하여 성능 저하를 초래한다.

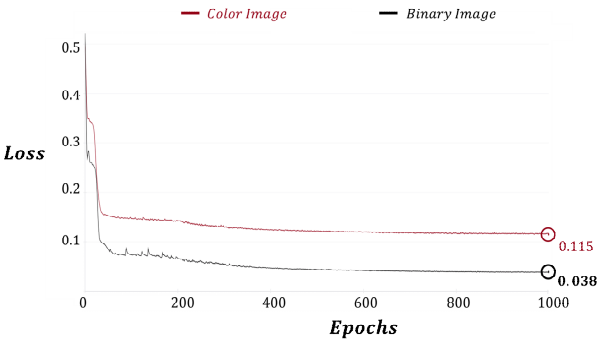

그림 7. MAE 학습에서 컬러 이미지와 흑백 이미지에 대한 최적화 비교
Fig. 7. Optimization comparison for color images and black and white images in MAE learning

실제로, 표 3에서 이진화가 족적 이미지 검색에 미치는 영향을 확인할 수 있다. FID-300 데이터 세트에서 이진화를 적용했을 때, Top 1%에서는 1%, Top 5%에서는 3.02%, 그리고 Top 10%에서는 3.54%의 성능 향상을 보였다. 유사하게, 경찰청이 제공한 데이터 세트에도 Top 1%에서는 1.42%, Top 5%에서는 4.09%, 그리고 Top 10%에서는 5.3%의 성능 향상을 기록하였다.

본 논문은 흑백 이미지에 대한 이진화를 넘어, 보다 많은 정보를 포함하고 있는 그레이 이미지의 변환에 대한 추가실험을 진행하였다. 흑백 이미지 이진화와 마찬가지로 OpenCV 모듈을 활용하였고 표 4는 이에 대한 비교 실험이다. 표 4에서 FID-300과 경찰청에서 제공한 데이터 세트 둘 다 성능에 큰 차이를 보이지 않았다. 이러한 결과는 흑백 이미지보다 많은 정보를 내장한 그레이 스케일 이미지가 족적 이미지 검색 분야에서는 유의미한 결과를 보이지 않음을 알 수 있다.

이와 같은 이유로 이미지 이진화는 3차원을 1차원으로 줄이면서 메모리의 효율성과 계산의 복잡도 이점을 제공할 뿐만 아니라, 최적화에도 도움을 주기 때문에 전반적인 성능 향상에 상당한 영향을 미침을 확인하였다.

표 3. 이미지 이진화에 따른 성능 비교
Table 3. Comparison of performance with image binarization

Dataset	Top 1%	Top 5%	Top 10%	Binarization
FID-300	72.8	80.08	87.74	✗
FID-300	**73.8**	**83.1**	**91.28**	✓
Police	58.88	64.02	69.2	✗
Police	**60.3**	**68.11**	**74.5**	✓

표 4. 흑백 이미지와 그레이 이미지 성능 비교
Table 4. Comparison of black and white and gray imaging performance

Dataset	Top 1%	Top 5%	Top 10%	Scale type
FID-300	**73.80**	83.10	**91.28**	Binary
FID-300	73.71	**83.12**	91.27	Gray
Police	**60.30**	**68.11**	**74.50**	Binary
Police	60.28	68.10	74.45	Gray

V. 결 론

본 논문에서는 이미지 자동 검색에서 방대한 경찰청 데이터를 활용하기 위해, 일반적인 접근 방식인 샴 네트워크에서 벗어나 자기 지도 학습인 MAE의 방식을 제안하였다. MAE로 이미지 패치들을 높은 비율로 랜덤하게 마스킹 처리하고, 나머지 낮은 비율의 이미지 패치들만으로 이미지를 복원하는 자기 지도 방식의 학습이 이루어져, 풍부한 정보를 포함한 특징 학습을 실현하였다. 아울러 이미지 이진화가 특징 학습에 대한 최적화를 가속시키는 것을 확인하였고, 이러한 방식으로 3개의 평가 지표 모두에서 기존의 방식들보다 높은 성능을 달성하였으며, 시각적인 결과에서도 기존의 방식보다 쿼리 이미지가 GT 참조 이미지에 잘 매칭되어 우수한 성능을 띄는 것을 확인하였다. 최종적으로, 방대한 데이터에 대한 라벨링 작업이 별도로 필요 없기 때문에 수많은 인적 자원과 비용을 절감하는 장점을 가져 미래 족적 이미지 검색에 대해 많은 기여를 할 것을 시사한다.

References

[1] I. Rida, L. Fei, H. Proenca, A. N. Ali, and A. Hadid, "Forensic shoe-print identification: a brief survey", ArXiv preprint ArXiv:1901.01431, Dec. 2019. https://doi.org/10.48550/arXiv.1901.01431.

[2] G. Alexandre, "Computerized classification of the shoeprints of burglars' soles", Forensic Science International, Vol. 82, No. 1, pp. 59-65, Sep. 1996. https://doi.org/10.1016/0379-0738(96)01967-6.

[3] P. D. Chazal, J. Flynn, and R. B. Reilly, "Automated processing of shoeprint images based on the Fourier transform for use in forensic science", IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 27, No. 3, pp. 341-350, Jan. 2005. https://doi.org/10.1109/TPAMI.2005.48.

[4] H. Kim and S. Kim, "Intelligent Non-contact Respiratory Monitoring System using CNN", Journal of KIIT, Vol. 22, No. 4, pp. 9-15, Apr. 2024. https://doi.org/10.14801/jkiit.2024.22.4.9.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need", in Proc. of the Conf. on Neural Information Processing Systems(NIPS), pp. 6000-6010, Dec. 2017. https://doi.org/10.48550/arXiv.1706.03762.

[6] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 10012-10022, Oct. 2021. https://doi.org/10.48550/arXiv.2103.14030.

[7] Z. Ma, Y. Ding, S. Wen, J. Xie, Y. Jin, Z. Si, and H. Wang, "Shoe-Print Image Retrieval With Multi-Part Weighted CNN", IEEE Access , Vol. 7, pp. 59728-59736, May 2019. https://doi.org/10.1109/ACCESS.2019.2914455.

[8] W. Liu and D. Xu, "Robust and Efficient Shoe Print Image Retrieval Using Spatial Transformer Network and Deep Hashing", in Proc. of the 4th

International Symposium on Signal Processing Systems (SSPS), pp. 89-95, Mar. 2022. https://doi.org/10.1145/3532342.3532356.

[9] D. Li, Y. Li, and Y. Liu, "Shoeprint Image Retrieval Based on Dual Attention Light HashNetwork", in Proc. of 4th International Conf. on Artificial Intelligence and Pattern Recognition(AIPR), Xiamen China, pp. 354-359, Sep. 2021. https://doi.org/10.1145/3488933.3488938.

[10] W. Yanjun, X. Wang, and T. Zhang, "Crime scene shoeprint retrieval using hybrid features and neighboring images", Information, Vol. 10, No. 2, pp. 324-331, Jan. 2019. https://doi.org/10.3390/info10020045.

[11] Y. Wu, X. Dong, G. Shi, X. Zhang, and C. Chen, "Crime Scene Shoeprint Image Retrieval: A Review", Science & Justice, Vol. 63, No. 4, pp. 439-450, Jul. 2023. https://doi.org/10.3390/electronics11162487.

[12] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese Neural Networks for One-shot Image Recognition", International Conf. on Machine Learning(ICML), Lille, France, Vol. 2, No. 1, Jul. 2015.

[13] G. Albert, J. Almazan, J. Revaud, and D. Larlus, "End-to-End Learning of Deep Visual Representations for Image Retrieval", IInternational Journal of Computer Vision, Vol. 124, No. 2, Jun. 2017. https://doi.org/10.48550/arXiv.1610.07940.

[14] J. Revaud, J. Almazan, R. S. Rezende, and C. R. Souza, "Learning with Average Precision: Training Image Retrieval with a Listwise Loss", in Proc. of the International Conf. on Computer Vision(ICCV), Seoul, Korea, pp. 5107-5116, Jun. 2019. https://doi.org/10.48550/arXiv.1906.07589.

[15] W. Su, X. Zhu, C. Tao, L. Lu, B. Li, G. Huang, Y. Qiao, X. Wang, J. Zhou, and J. Dai, "Towards All-in-one Pre-training via Maximizing Multi-modal Mutual Information", in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition(CVPR), Seattle WA, USA, pp. 15888-15899, Jun. 2023. https://doi.org/10.48550/arXiv.2211.09807.

[16] P. Wang, S. Wang, J. Lin, S. Bai, X. Zhou, J. Zhou, X. Wang, and C. Zhou, "One-Peace: Exploring One General Representation Model Toward Unlimited Modalities", ArXiv preprint ArXiv:2305.11172, May 2023. https://doi.org/10.48550/arXiv.2305.11172.

[17] S. Srivastava and G. Sharma, "OmniVec: Learning robust representations with cross modal sharing", in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition(CVPR), Waikoloa, Hawaii, pp. 5484-5494, Jun. 2024. https://doi.org/10.48550/arXiv.2311.05709.

[18] K. He, X. Chen, S. Xie, P. Dollar, and R. Cirshick, "Masked Autoencoders Are Scalable Vision Learners", in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition(CVPR), New Orleans, Louisiana, pp. 16000-16009, Jun. 2022. https://doi.org/10.48550/arXiv.2111.06377.

[19] M. Singh, Q. Duval, K. V. Alwala, H. Fan, V. Aggarwal, A. Adcock, A. Joulin, P. Dollar, C. Feichtenhofer, R. Girshick. R. Girdhar, and I. Misra, "The effectiveness of MAE pre-pretraining for billion-scale pretraining", in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition(CVPR), Paris France, pp. 5484-5494, Jun. 2023. https://doi.org/10.48550/arXiv.2303.13496.

[20] J. Dia, H. Qi, Y. X. Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable Convolutional Networks", in Proc. of the IEEE Conf. on Vision (ICCV), Venice, Italy, pp. 764-773, Jul. 2017. https://doi.org/10.48550/arXiv.1703.06211.

[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", ArXiv preprint ArXiv:2010.11929, Oct. 2020. https://doi.org/10.48550/arXiv.2010.11929.

[22] Y. Li, K. Zhang, J. Cao, R. Timofte, and L. V.

Gool, "LocalViT: Bringing Locality to Vision Transformers", ArXiv preprint ArXiv:2104.05707, Apr. 2021. https://doi.org/10.48550/arXiv.2104.05707.

[23] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for Semantic Segmentation", in Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV), pp. 7262-7272, Jun. 2021. https://doi.org/10.48550/arXiv.2105.05633.

[24] G. Bradski, "The OpenCV library", Dr. Dobb's Journal: Software Tools for the Professional Programmer , Vol. 25, No. 11, pp. 120-125, Nov. 2000.

[25] S. Shafique, B. Kong, S. Kong, and C. Fowlkes, "Creating a Forensic Database of Shoeprints From Online Shoe-Tread Photos", in Proc. of the IEEE Winter C. on Applications of Computer Vision(WACV), Waikoloa, Hawaii, Jan. 2023. https://doi.org/10.48550/arXiv.2205.02361.

저자소개

이 창 엽 (Chang-Yeop Lee)

2023년 : 인제대학교 컴퓨터공학부(공학사)
2022년 ~ 현재 : 포항공과대학교 인공지능연구원 연구부 연구원
관심분야 : Object detection, Monocular depth estimation, Semantic segmentation

김 동 주 (Dong-Ju Kim)

2010년 : 성균관대학교 전기전자컴퓨터공학과(공학박사)
2011년 ~ 2015년 : 대구 경북 과학기술원 IT 융합 연구부 선임연구원
2015년 ~ 2015년 : 동서대학교 컴퓨터공학부 조교수
2016년 ~ 현재 : 포항공과대학교 인공지능연구원 연구부 연구부장
관심분야 : Computer vision, Face recognition, Deep learning

서 영 주 (Young-Joo Suh)

1996년 : Georgia Institute of Technology 컴퓨터공학(공학박사)
1988년 ~ 현재 : 포항공과대학교 컴퓨터공학과 교수
2016년 ~ 2020년 포항공과대학교 정보통신연구소 소장
관심분야 : AI, IoT, Action recognition, Indoor positioning

황 도 경 (Do-Kyung Hwang)

2020년 ~ 2020년 : 포항공과대학교 PMC 연구실 연구원
2020년 : 부산대학교 전자공학과(공학석사)
2020년 ~ 현재 : 포항공과대학교 인공지능연구원 연구부 연구원
관심분야 : Data restoration, DL/ML, Image processing, Signal processing