

Contactless User Identification System using Multi-channel Palm Images Facilitated by Triple Attention U-Net and CNN Classifier Ensemble Models

Inki Kim*, Beomjun Kim*, Sunghee Woo**, Jeonghwan Gwak***

*Student, Dept. of IT-Energy Convergence, Korea National University of Transportation, Chungju, Korea

*Student, Dept. of IT-Energy Convergence, Korea National University of Transportation, Chungju, Korea

**Professor, Dept. of Computer Engineering, Korea National University of Transportation, Chungju, Korea

***Professor, Dept. of Software, Korea National University of Transportation, Chungju, Korea

[Abstract]

In this paper, we propose an ensemble model facilitated by multi-channel palm images with attention U-Net models and pretrained convolutional neural networks (CNNs) for establishing a contactless palm-based user identification system using conventional inexpensive camera sensors. Attention U-Net models are used to extract the areas of interest including hands (i.e., with fingers), palms (i.e., without fingers) and palm lines, which are combined to generate three channels being ped into the ensemble classifier. Then, the proposed palm information-based user identification system predicts the class using the classifier ensemble with three outperforming pre-trained CNN models. The proposed model demonstrates that the proposed model could achieve the classification accuracy, precision, recall, F1-score of 98.60%, 98.61%, 98.61%, 98.61% respectively, which indicate that the proposed model is effective even though we are using very cheap and inexpensive image sensors. We believe that in this COVID-19 pandemic circumstances, the proposed palm-based contactless user identification system can be an alternative, with high safety and reliability, compared with currently overwhelming contact-based systems.

▶ **Key words:** Palm-based Identification, Contactless Identification System, Multi-channel image, Attention U-Net, Ensemble of Pre-trained CNN Models

[요 약]

본 논문에서는 기존의 스마트폰 카메라 센서를 사용하여 비접촉식 손바닥 기반 사용자 식별 시스템을 구축하기 위해 Attention U-Net 모델과 사전 훈련된 컨볼루션 신경망(CNN)이 있는 다채널 손바닥 이미지를 이용한 앙상블 모델을 제안한다. Attention U-Net 모델은 손바닥(손가락 포함), 손바닥(손바닥 미포함) 및 손금을 포함한 관심 영역을 추출하는 데 사용되며, 이는 앙상블 분류기로 입력되는 멀티채널 이미지를 생성하기 위해 결합 된다. 생성된 데이터는 제안된 손바닥 정보 기반 사용자 식별 시스템에 입력되며 사전 훈련된 CNN 모델 3개를 앙상블 한 분류기를 사용하여 클래스를 예측한다. 제안된 모델은 각각 98.60%, 98.61%, 98.61%, 98.61%의 분류 정확도, 정밀도, 재현율, F1-Score를 달성할 수 있음을 입증하며, 이는 저렴한 이미지 센서를 사용하고 있음에도 불구하고 제안된 모델이 효과적이라는 것을 나타낸다. 본 논문에서 제안하는 모델은 COVID-19 팬데믹 상황에서 기존 시스템에 비하여 높은 안전성과 신뢰성으로 대안이 될 수 있다.

▶ **주제어:** 손바닥 기반 신원 인식, 비접촉식 신원 인식 시스템, 멀티채널 영상, 어텐션 유넷, CNN 모델 앙상블

- First Author: Inki Kim, Corresponding Author: Jeonghwan Gwak
- *Inki Kim (cv2@kakao.com), Dept. of IT-Energy Convergence, Korea National University of Transportation
- *Beomjun Kim (zhskfkseh@gmail.com), Dept. of IT-Energy Convergence, Korea National University of Transportation
- **Sunghee Woo (shwoo@ut.ac.kr), Dept. of Computer Engineering, Korea National University of Transportation
- ***Jeonghwan Gwak (jgwak@ut.ac.kr), Dept. of Software, Korea National University of Transportation
- Received: 2022. 02. 21, Revised: 2022. 02. 24, Accepted: 2022. 03. 18.

I. Introduction

현대 사회에서는 모바일 기기, 데스크톱, 보안 시설 및 보안 관리 시스템 등 보안 시스템의 적용이 필요한 경우에 다양한 사용자 인식 방법을 사용해 사용자에게 맞는 적절한 권한을 부여하는 방법을 사용하고 있다. 하지만 패스워드[1], 주민등록번호는 유출의 위험이 크며 보안성 또한 확보되지 않아 이중 인증을 사용하는 경우가 많다. 이중 인식의 경우 2차 비밀번호를 입력하거나, 특별히 제작된 USB[2]를 인식시키는 등 다양한 대안이 나타났다. 하지만 이 또한 분실과 복제에 대한 위험이 존재하며, 유출될 수 있다는 단점이 존재한다. 이에 대한 대안으로 생체정보를 활용한 사용자 인식 방법이 사용된다. 생체 인식은 사람마다 가지고 있는 고유한 생체정보를 이용해 개인을 식별하거나 인증하는 정보 보안 기술이다. 분실 및 위/변조의 위험이 낮고, 높은 신뢰성 및 편의성을 제공하여 주목받고 있다. 생체 인식에는 신체적 생체 인식과[3], 습관적 생체 인식[4] 두 가지 종류가 존재한다. 전자인 신체적 생체 인식은 사람의 신체 구조, 유전자를 활용한 생물학적인 측정 방법이다. 일반적으로 알려진 지문[5], 정맥[6], 홍채[7], 얼굴 인식[8] 등이 이에 포함된다. 현재 모바일 기기에서 가장 많이 사용되고 있는 지문인식의 경우에는, 센서의 값이 저렴하다는 가장 큰 장점이 있다. 하지만 손가락에 상처가 나거나, 변형이 생길 수 있다는 단점이 존재하며 이물질이나 오염도에 따라 성능에 크게 영향을 미친다. 정맥, 홍채 인식의 경우에는 보안성이 좋다는 장점이 있지만, 시스템 설치 비용이 매우 비싸다는 단점이 존재한다. 얼굴 인식의 경우는 주변광에 따라 인식 성능이 크게 차이가 나며, 현재 COVID-19 팬데믹이 시작됨에 따라 마스크를 착용하기 시작하였는데, 마스크를 착용한 사람의 얼굴 인식이 아직 완벽히 보완되지 않았다. 후자의 경우인 습관적 생체 인식의 경우 생물학적인 측정을 하지 않고도 신원을 확인하는 방법으로 다중 객체추적 알고리즘[9, 10] 기반 걸음걸이, 서명, 음성인식 등이 습관적 생체 인식의 경우에 속한다. 하지만 복제가 어렵지 않고, 권한이 있는 사용자인지 추가로 확인할 방법이 없다는 단점이 존재한다.

또한, COVID-19 팬데믹이 시작되면서 대중적으로 사용되는 지문인식의 경우 전파감염의 위험성이 대두되면서 비접촉/비대면 방식의 사용자 인식이 유행되기 시작하였다. 이에 따라 본 논문에서는 기존에 사용되던 생체 인식 기반 사용자 인식 방법에서, 비접촉/비대면으로 활용될 수 있고, 비접촉 생체 인식으로서 보편성, 유일성, 영구성, 획득성, 정확성 등을 가지고 있는 손금을 기반으로 하는

생체 인식 방법을 제안한다. 정보통신정책연구원(KISDI)에서는 생체 인식 기술에 활용되기 위해 고유한 특성을 가질 필요가 있다고 주장한다[11]. 첫 번째로 보편성은 모든 사람이 가지고 있는 생체 특성이어야 한다. 두 번째로 유일성은 같은 특성을 가진 사람이 없어야 한다. 세 번째로 영구성은 절대 변화하거나 변경되지 않아야 한다. 네 번째로 획득성은 센서로부터 생체 특성 정보의 추출 및 정량화가 용이해야 한다. 마지막으로 정확성은 시스템의 정확도, 처리속도, 내구성 등을 이야기한다. 본 논문에서 제안하는 생체 인식 기반 손금을 이용한 신원 인식 방법은 위에서 언급한 생체 인식 기술에 활용되기 위한 고유한 특성을 모두 가지고 있다. 또한, 상처에 민감하거나 오염도에 따라 정확도가 크게 변하는 지문과는 달리, 손바닥 전체의 영역을 이용하여 생체 인식을 하므로 상처나 오염도에 따라 성능에 영향을 미치지 않는다. 또한, 지문의 특징은 땀샘이 돌기 돼 튀어나온 융선(ridge), 융선과 융선 사이 오목한 부분인 골(valley)로 구분된다.

본 논문에서 제안하는 손바닥의 핵심 특징들을 추출하는 방법은 손가락이 포함된 손바닥의 영역, 손가락이 제외된 손바닥의 영역, 손금 영역으로 구성된 멀티채널 이미지를 통해 손가락이 포함된 손바닥의 영역부터 차례대로 Attention을 강조할 수 있다. 손가락이 포함된 손바닥의 영역에서는 가장 크게 두드러지는 손가락의 길이, 손바닥의 크기 등 전체적인 윤곽선이나 길이의 특징을 포함하며, 손가락이 제외된 손바닥의 영역에서는 손바닥의 특징과 손금의 영역을 포함해 멀티채널 이미지의 마지막 구성요소인 손금 영역의 이미지에서 손금의 영역을 강조시키는 역할을 할 수 있다. 이로 인해 단순히 인공지능 모델이 손금만을 추출하였을 때, 손금의 추출이 성공적으로 되지 않았다고 하더라도 손가락이 제외된 손바닥의 영역, 손가락이 포함된 손바닥의 영역에서 개인이 가지고 있는 손바닥의 고유한 특징을 통해 신원 인식을 도와줄 수 있다는 장점이 있다.

II. Preliminaries

1. Related works

국내에서는 생체 인식에 관한 연구는 활발하게 이루어지고 있지만, 손금의 특징을 이용한 생체 인식 연구는 비교적 활발하게 연구가 되지 않고 있다. 컴퓨터 비전을 통한 손금 인식부터, 딥러닝을 이용한 손금과 정맥을 이용한 사용자 인식의 방법이 다양하게 적용되었다. Fig. 1은 국

내 논문[12]에서 제안된 손금 및 손바닥 정맥 기반 신원 인식 아키텍처이다.

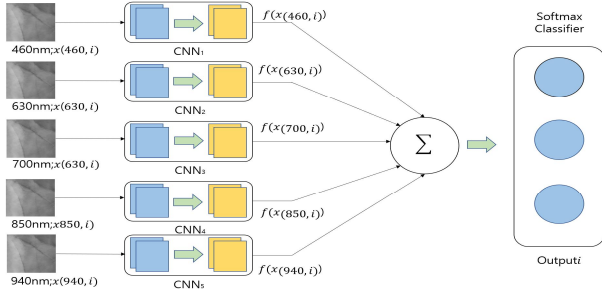


Fig. 1. Palmline and Palm Vein Based Identification Model[12]

김슬빈 외[12]는 손바닥 영상에서 관심 영역을 검출하고 해당 영역을 기반으로 사용자를 인식하는 심층 신경망 기반 사용자 인식 시스템이다. 획득한 영상에 대해 Otsu Thresholding을 수행하고, 전경 영역에서 Local Minimal을 기반으로 ROI를 추출한다. 획득한 이미지는 화질 열화가 발생하기 쉬운 근적외선 카메라로 촬영이 되어, Contrast Limited Adaptive Histogram Equalization을 적용하고, High Boost를 통해 이미지 화질을 개선한다. 해당 이미지를 심층 신경망의 앙상블을 이용해 사용자를 인식하는 기법이다. 하지만 해당 논문은 본 논문에서 제안하는 손가락의 손금 정보보다는 정맥을 통해 사용자를 인식하는 기법이다. 국내에서는 손금과 답러닝을 통해 사용자를 인식하는 방법에 관한 연구는 거의 없으며, 멀티채널을 통해 새롭게 이미지를 구성하고 사용자를 인식하는 방법 또한 제안된 바가 없다. With Finger를 통해 손바닥 전체의 영역에 Attention을 부여하고, Without Finger를 통해 손가락이 제외된 손바닥 영역에 추가적인 Attention을 부여하고, Palm Line을 통해 핵심적인 손금 영역에 Attention을 부여하는 피라미드 방식의 Attention을 통하여 생체 인식에 사용할 수 있는 정보들에 Attention 할 수 있게 한다.

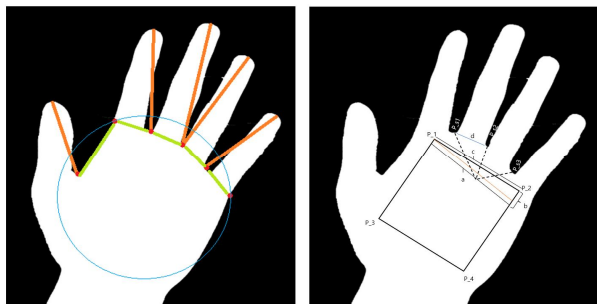


Fig. 2. Calculated Palm Area and Points for ROI Extraction[13]

해외에서는 Fig. 2와 같이 Palm Print[13]와 Hand Geometry[13]에 Morphology, Template-based Matching, Key Point Detection 등을 이용해 사진에 필터링 작업을 진행하여 일치하는 점의 위치, Geometry Data 등을 이용해 User Identification을 적용한 바가 있다. 즉, 컴퓨터 비전 Algorithm을 이용하여 손바닥 특징에 접근해 등록된 사용자의 사진과 비교하여 신원 인식을 진행한다. 하지만 이러한 방법은 입력된 사진의 조도, 배경 등 주변 환경에 의해 복잡도가 상승할수록 보안성과 정확도가 떨어진다는 단점이 존재한다. 예를 들어 손바닥의 점이 주변 조도에 의하여 사진상에 제대로 투영되지 않는 경우 전혀 다른 사람으로 예측하거나 등록되지 않은 사용자로 인식할 확률이 높다. 또한, 손바닥만을 포함한 Polygon 형태의 관심 영역(ROI)을 설정할 때 배경이 복잡하거나 손바닥 색과 비슷하다면 정확하게 관심 영역의 추출이 어려울 것이다. 이러한 문제점들을 해결하기 위해 비교를 위해 저장되는 샘플 이미지의 개수를 늘린다면, 이는 단 하나의 사진에서 최악의 경우 모든 이미지를 비교해야 하는 Method 특성상 연산량이 매우 많아진다. 또한, 다중 Classification을 위해 각 클래스에 해당하는 사람들의 샘플 이미지를 수집하는 경우에도 같은 문제가 발생한다. 다른 방법으로는 SIFT, TCCM의 융합 데이터를 이용하여 SVM(Support Vector Machine) 모델에 입력하여 신원 인식을 진행하는 연구[14]도 있다. 하지만 L. G. Oldal and A. Kovács[13]과 비슷하게 특징점을 추출하는 과정에서 입력된 사진의 조도, 배경 등 주변 환경의 영향에 따라 제한을 받을 수 있으며, 이는 보안성과 정확도가 떨어진다는 단점이 존재한다. 이는 관심 영역의 추출과 연산량 증가의 문제점 또한 여전히 존재하며 신원 인식 시스템은 보통 건물의 출입구에 설치하기 때문에 주변 환경이 복잡하며 일시에 따라 조도 또한 수시로 바뀌기 때문에 추가적인 조치가 필요해 시스템 설치 비용에 대한 문제점을 해결하기에 부족하다. 추가적으로 단일 모델을 사용하여 Attention U-Net을 변형한 Context Fusion[15]을 이용한 방법, 추출한 ROI에서 Feature Extraction[16]을 이용한 방법 등이 있다. 본 논문에서 제안하는 방법은 위에서 설명한 문제점들을 최소화하기 위해 Segmentation을 통한 관심 영역(ROI) 설정, 관심 영역을 이용한 특징점 추출, 복합 비교를 위한 멀티채널 이미지 생성과 컴퓨터 비전 알고리즘을 이용하여 복잡한 배경에서도 정확하게 손바닥만을 관심 영역을 설정하고 손금 등의 특징점을 추출하여 신원 인식을 위한 데이터 재구성/생성을 통해 보안성과 정확도 또한 높이고, 시스템 설치 비용을 최소화할 수 있는 연구를 진행하였다.

III. Proposed Method

1. Image Segmentation

손금의 핵심 특징을 추출하기 위해 본 논문에서 제안한 멀티채널 이미지를 생성하기 위해서는, 손가락이 포함된 손바닥 영역, 손가락이 제외된 손바닥 영역, 손금 영역의 이미지가 필요하다. 하지만 사람이 매뉴얼 하게 Ground truth를 제작하기에는 너무 많은 노동력과 시간이 필요하다. 따라서 본 논문에서는 이미지 분할(Image Segmentation)을 목적으로 제안된 Attention U-Net[17]을 사용한다. 기존의 U-Net[18]은 Biomedical 분야에서 세포(Cell)의 분할(Segmentation)을 목적으로 제안된 Fully-Convolutional Network 모델이다. 기존의 U-Net은 Contracting Path와 Expanding Path를 통해 이미지의 Context 정보를 얻고, 지역화(Localization)를 위한 구조가 대칭 형태로 구성되어 있다. Fig. 3은 U-Net의 아키텍처이다.

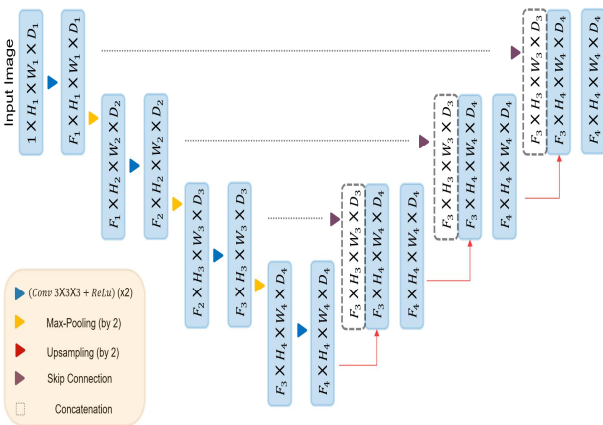


Fig. 3. U-Net Architecture[18]

이에 반해 본 논문에서 사용한 Attention U-Net은 U-Net에서 추출된 저차원의 특징들을 고차원으로 변환하는 과정에서 관심 영역을 강조할 수 있는 Attention Gate를 이용해 관심 영역이 아닌 부분은 가중치를 조정하여 특징에서 중요한 영역을 강조할 수 있으므로, 손바닥 신원 인식에 활용할 수 있는 핵심 정보를 더 잘 나타낼 수 있도록 한 네트워크이다.

기존의 U-Net에서 Skip-connection, Upsampling 과정에서 Attention gate와 Gating signal이 적용된 것이 가장 큰 차이점이다. Fig. 4는 Attention U-Net의 아키텍처이며, Fig. 5는 Attention Gate의 구조이다. Gating Signal의 G 는 직전의 레이어에서 가져오는 특징 정보이며, Skip-Connection에서 가져오는 X 보다는 Feature Representation 능력이 좋다. 반면에 X 는 Skip-Connection에서 전해지며, G 보다는 Feature

Representation 능력은 떨어지지만, 더 많은 공간 정보를 포함하고 있다.

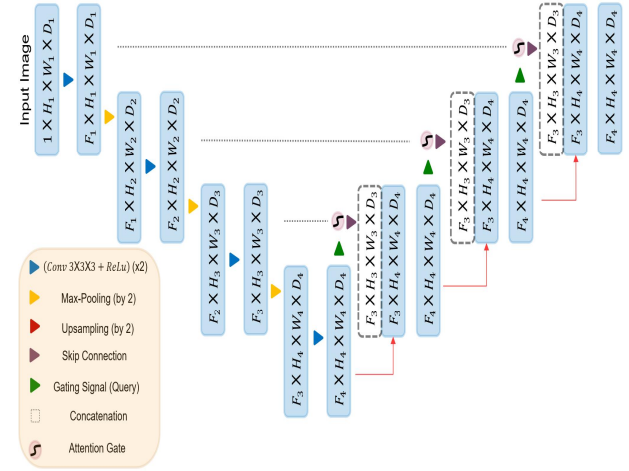


Fig. 4. Attention U-Net Architecture[17]

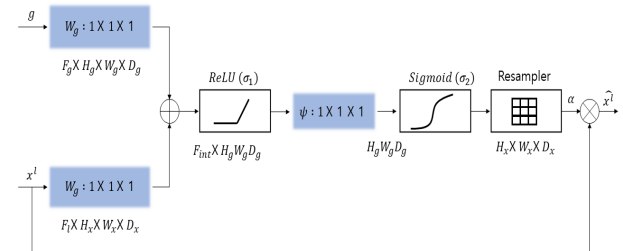


Fig. 5. Attention Gate Architecture[17]

2. Dataset

멀티채널 이미지를 구성하기 위해서는, 먼저 손가락이 포함된 손바닥 영역, 손가락이 제외된 손바닥 영역, 손금 영역의 이진화(Binary)된 이미지가 필요하다. 따라서 세 가지 분야의 Image Segmentation 네트워크를 학습하고, 예측한 결과를 멀티채널 이미지로 구성한다. Table 1은 학습에 사용된 이미지의 수를 나타낸다.

Table 1. Dataset Distribution

Dataset	Training	Validation	Test
With Finger	3,292	412	411
Without Finger	1,185	395	396
Palm Line	1,095	138	136

데이터셋의 모든 이미지는 각기 다른 20명의 손바닥을 스마트폰 카메라로 촬영한 이미지이다. 촬영에 사용된 휴대폰은 Samsung Galaxy S21 Ultra, iPhone XR이다. 두 스마트폰 모두 촬영되는 이미지의 크기가 다르기 때문에, (224 X 224)로 이미지의 크기를 동일하게 변경하여 학습에

사용하였다. 원본 이미지는 988장으로, augmentations 라이브러리를 이용하여 데이터 증강(Data augmentation)을 적용하거나 여러 장의 배경 이미지와 학습 시에 사용하고자 생성한 Ground truth 이미지를 이용하여 복잡한 배경에서도 정확하게 Segmentation이 진행될 수 있도록 데이터 증강 기법을 적용하였다. 손금 영역의 이미지는 Ground truth가 너무 얇으므로 이미지를 Downsampling할 경우, 이진화 이미지의 픽셀들의 정보가 무의미한 값들로 변하기 때문에 모폴로지의 팽창(Morphology dilation) 연산을 통해 Ground truth의 손금을 조금 두껍게 변환하였다. 또한, 학습 데이터로 사용되는 영상 프레임의 경우 미세한 배경의 움직임, 의미 없는 픽셀은 전처리 과정[19] 및 영상 복잡도 분석[20]을 통하여 효과적으로 제거되어야 하며, 손금 영역에서는 이미지 자체에 Attention을 부여하기 위해 손바닥 포함 영역 Image segmentation 결과를 Bitwise AND 연산을 통해 배경을 제거하였다. Fig. 6는 3개의 데이터셋의 샘플을 나타낸다.

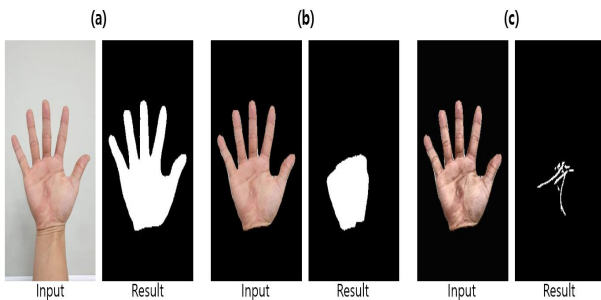


Fig. 6. (a) With Finger Data Sample (b) Without Finger Data Sample (c) Palm Line Dataset Sample

데이터셋을 각각 확인해보면 손가락을 포함한 손바닥 Segmentation Model의 학습 시 입력되는 이미지들은 손바닥 사진을 촬영한 원본 이미지이다. 원본 이미지가 입력되면 손바닥 모양의 단일 채널 마스크 이미지가 생성되는데 흰색 부분이 손바닥, 검은색 부분이 손바닥을 제외한 배경으로 표현된다. 손가락을 제외한 손바닥 Segmentation Model의 학습 시 입력되는 이미지들은 손가락을 포함한 손바닥 Segmentation Model의 예측 결과와 원본 이미지를 Bitwise AND 연산을 통해 배경을 제거한 이미지이다. 해당 이미지가 입력되면 손가락을 제외한 손바닥 부분만이 포함된 단일 채널 마스크 이미지가 생성되며 흰색 부분이 손가락을 제외한 손바닥, 검은색 부분이 손가락을 표현한다. 마지막으로 손금 Segmentation Model은 손가락을 제외한 손바닥 Segmentation Model에 입력되는 이미지에 CLAHE를 이용하여 이미지에 굴곡 부분을 강조하여 손금 부분이 강조된 이미지를 생성하고 학습 시 사용된다.

3. Image Segmentation Experiments

학습과 실험에 사용된 환경은 Table 2와 같다.

Table 2. System Environment

Item	Value
CPU	Intel i9-10900K
Memory	64GB
GPU	NVIDIA GEFORCE RTX 3090
OS	Windows, Ubuntu Server 18.04 LTS
Python Ver	3.9.6
PyTorch Ver	1.9.1+cu11.1

하이퍼파라미터(Hyperparameters)의 경우 lr(Learning Rate)은 0.0001에서 0.001 까지, Loss Function은 BCE(Binary Cross Entropy Loss)와 BCEWL(Binary Cross Entropy With Logits Loss) 두 가지로 실험하였으며 각각 100회씩 훈련을 진행하였다. 그중 IoU Score 기준 가장 높은 점수가 나타나는 Checkpoint를 사용하였으며, 각 데이터셋마다 학습에 사용된 하이퍼파라미터는 Table 3과 같다.

Table 3. Attention U-Net Hyperparameters

Dataset	learning rate	batch size	loss function	Optimizer	Epoch
With Finger	0.0005	16	BCE WL	Adam	100
Without Finger	0.0005	16	BCE WL	Adam	100
Palm Line	0.001	16	BCE	Adam	100

네트워크를 학습할 때, 각 Epoch에서 가장 높은 IoU(Intersection over Union) Score를 나타내는 Epoch의 Checkpoint를 사용하였다. Fig. 7은 With Finger 데이터셋으로 학습한 네트워크와 Without Finger 데이터셋으로 학습한 네트워크, Palm Line 데이터셋으로 학습한 네트워크의 입력 이미지에 대한 예측 결과를 나타낸다. 입력 이미지는 학습에 사용된 이미지가 아닌, 테스트 단계에서 사용되는 테스트 이미지를 사용하였다. Fig. 8과 Table 4는 각 모델의 Loss 훈련 지표와 Best IoU Score를 보여준다.

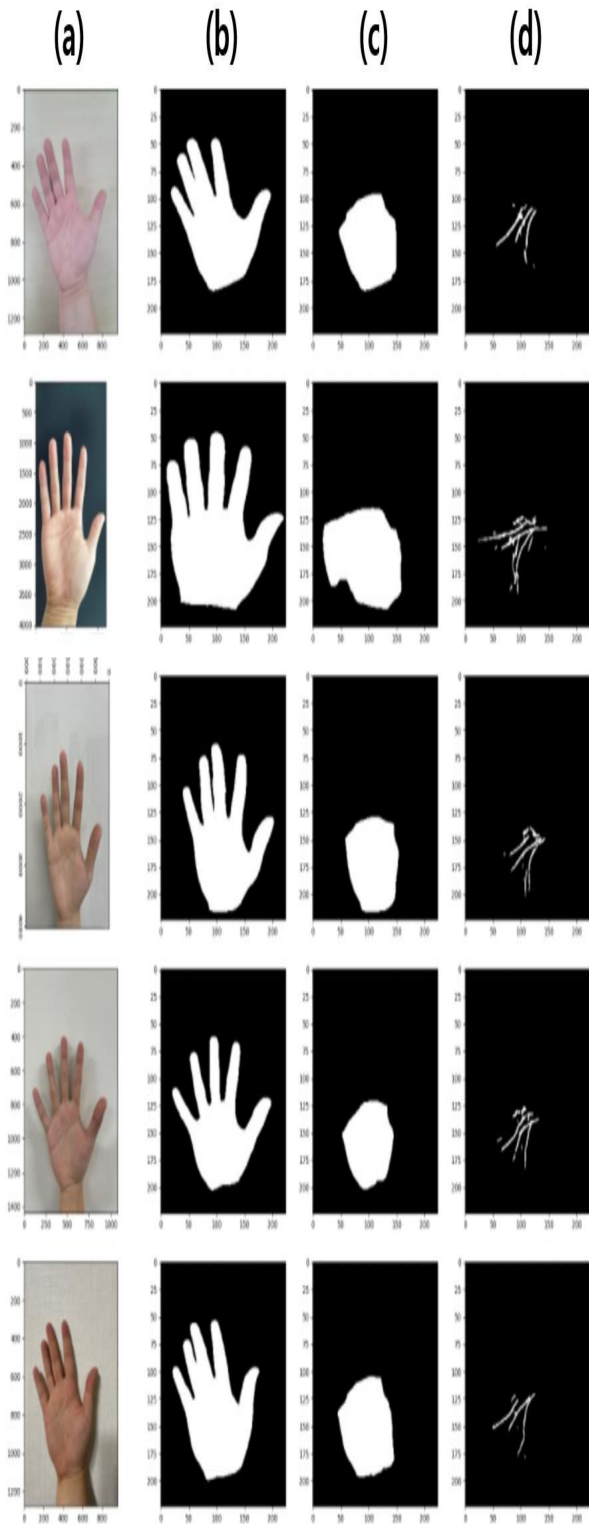


Fig. 7. (a) Input (b) Hands (with fingers) prediction results (c) Palm (without fingers) prediction results (d) Palm line prediction results

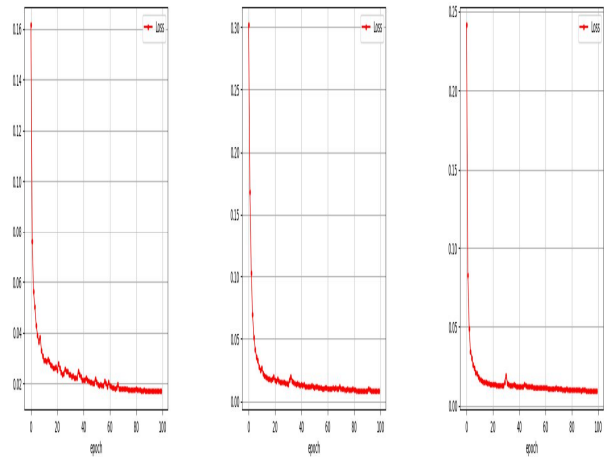


Fig. 8. (a) Hands (with fingers) (b) Palm (without fingers) (c) Palm line

Table 4. Best IoU Score

	With Finger	Without Finger	Palm Line
Best Score	0.950398407	0.952179392	0.59666117

Fig. 7에서 보인 결과들처럼 예측된 이미지를 원본 이미지와 Bitwise AND 연산을 통해 배경을 제거하고, 순차적으로 손금의 영역에 Attention 할 수 있는 멀티채널 이미지를 생성한다. Fig. 9는 Bitwise AND 연산을 이용한 결과 예시이다.

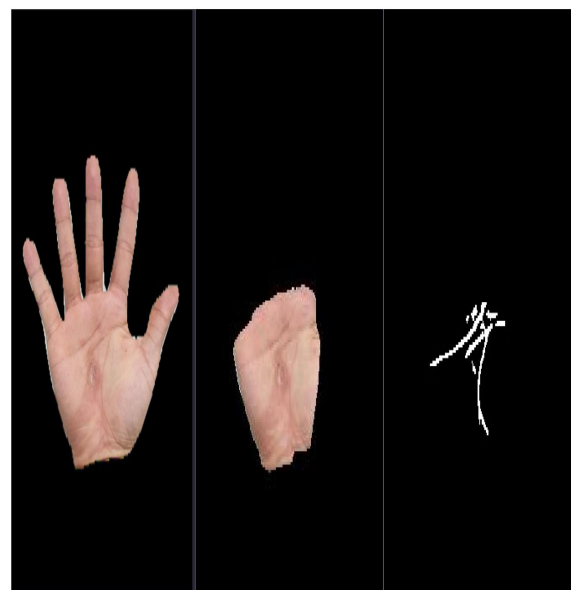


Fig. 9. Results of Bitwise AND Operation

4. Recognition using Multi-channel images

3가지 데이터셋으로 학습된 네트워크들의 예측 결과는, 한 사람의 손바닥에서 3개의 이미지로 나타나며 이는 멀티 채널 이미지의 각 채널 인덱스 요소가 된다. 일반적으로 이미지는 RGB 또는 BGR처럼 3개의 채널을 가지는 이미지이다. 즉, 본 논문에서 언급하는 Multi-channel Image는 각 네트워크의 예측 결과인 이미지를 3차원 이미지에서의 인덱스로 구성한 새로운 이미지 데이터로 정의한다. 손가락 포함 손바닥 네트워크는 1차원, 손가락 제외 손바닥 네트워크는 2차원, 손금 영역 이미지는 3차원으로 구성하여 기존의 이미지와는 다른 멀티채널 이미지를 생성한다. Fig. 10은 멀티채널 이미지의 각 인덱스 별 요소를 구성하는 방법을 나타낸다. 1차원 인덱스에서는 전체 손바닥의 영역의 정보를 포함하며, 2차원 인덱스에서는 주요 손금들의 위치 정보를 제공한다. 마지막으로 3차원 인덱스에서는 생체 인식에 핵심이 되는 손금을 나타내며, 1차원의 전체 손바닥 영역에서 3차원의 손금 영역까지 점진적으로 축소되는 과정에서 핵심 특징들에 대해서 Attention을 부여하는 효과를 가질 수 있다.

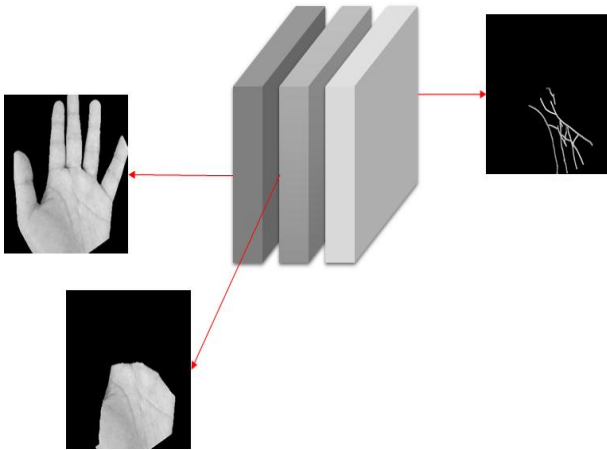


Fig. 10. Components for Constructing a Multi-channel image

위 과정을 통해 사용자 한 명의 신원 인식을 통한 멀티 채널 이미지를 구성할 수 있다. 본 논문에서는 7명에 대한 신원 인식을 위해 1,793개의 이미지를 수집/증강하였다. 생성된 멀티채널 이미지는 사전에 학습된 CNN을 통해 학습하고, 가장 성능이 높은 3개의 CNN을 사용하여 Attention U-Net을 통해 생성한 멀티채널 이미지를 통해 신원 인식을 위한 분류 모델로 만들 수 있다. Table 5는 사전에 학습된 CNN을 통해 멀티채널 이미지를 분류하는 성능을 나타낸 테이블이다. Fig. 11은 Attention U-Net을 통해 손바닥 Feature들을 Segmentation하고, 멀티채널 이미지를 생

성하는 네트워크를 나타내며, Fig. 12는 멀티채널 이미지의 Feature를 추출하고 Classification을 통해 신원 인식을 진행하는 네트워크를 나타낸다.

Table 5. Pre-trained CNN Classification Performance

Model	Loss func	Loss Score	epoch
AlexNet[21]	BCE	7.4078	30
DenseNet121[22]	BCE	1.4184	30
DenseNet169[22]	BCE	0.8247	30
EfficientNet_B0[23]	BCE	0.3759	30
Inception_v3[24]	BCE	4.9198	30
MnasNet[25]	BCE	0.4005	30
MobileNet[26]	BCE	2.5728	30
ResNet50[27]	BCE	8.1695	30
ResNet101[27]	BCE	6.9537	30
Resnext50[28]	BCE	2.8696	30
Resnext101[28]	BCE	4.4661	30
ShuffleNet[29]	BCE	2.1444	30
VGG16[30]	BCE	1.6985	30
VGG19[30]	BCE	5.0885	30
Ensemble	BCE	0.2338	30

사전 학습된 CNN 분류 모델을 단일로 사용하여 Feature를 추출하고 신원 인식 모델에 통과 시켜 훈련을 진행하였다. 사용된 모델은 AlexNet, DenseNet121, DenseNet 169, EfficientNet_B0, Inception_v3, MnasNet, MobileNet, ResNet50, ResNet101, Resnext50, Resnext101, ShuffleNet, VGG16, VGG19를 이용하였으며 훈련 반복 횟수는 30번을 기준으로 BCE(Binary Cross Entropy) Loss 함수를 이용하여 Top-3(가장 낮은 Loss Score를 가진 3개의 모델)을 선정하고 앙상블 기법[31]을 이용하여 앙상블 모델 또한 같은 조건으로 실험을 진행하였다. 실험 결과는 본 연구에서 사용된 데이터셋을 기준으로 단일 모델은 3위로 DenseNet169가 0.8247, 2위로 MnasNet이 0.4005 그리고 1위로 EfficientNet_B0가 0.3759의 Loss Score로 Top-3모델로 선정되었으며 위 3가지 모델을 앙상블 하여 실험한 결과로는 Loss Score 0.2338로 최고 성능을 가진 모델로 선정되었다. 따라서 앙상블 모델을 이용한 신원 인식 시스템을 설계하였다.

신원 인식을 위한 Classification 모델의 하이퍼 파라미터는 Table 6과 같다.

Table 6. Classification Model Hyperparameters

learning rate	batch size	loss function	Optimizer	Epoch
0.001	16	CE	Adam	30

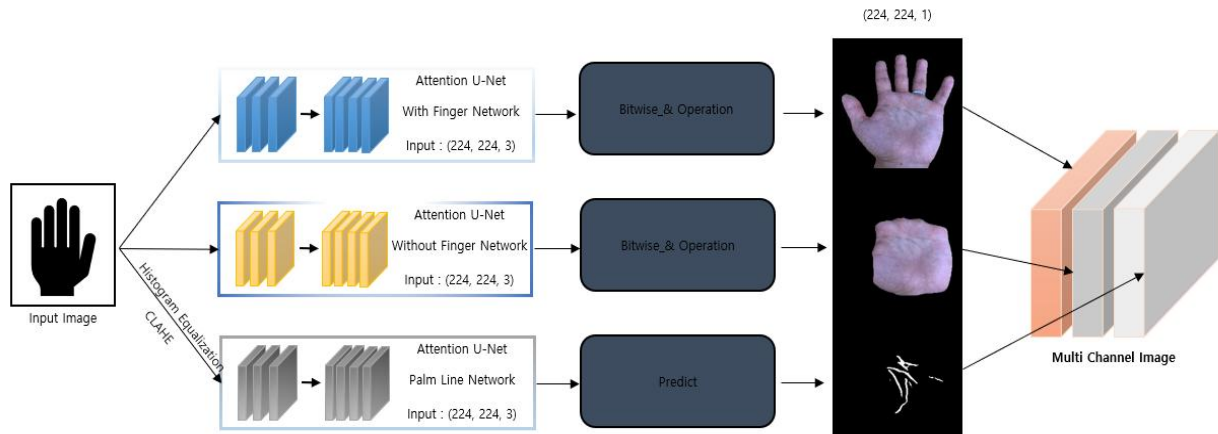


Fig. 11. Multi-channel image Creation Network Based on Attention U-Net

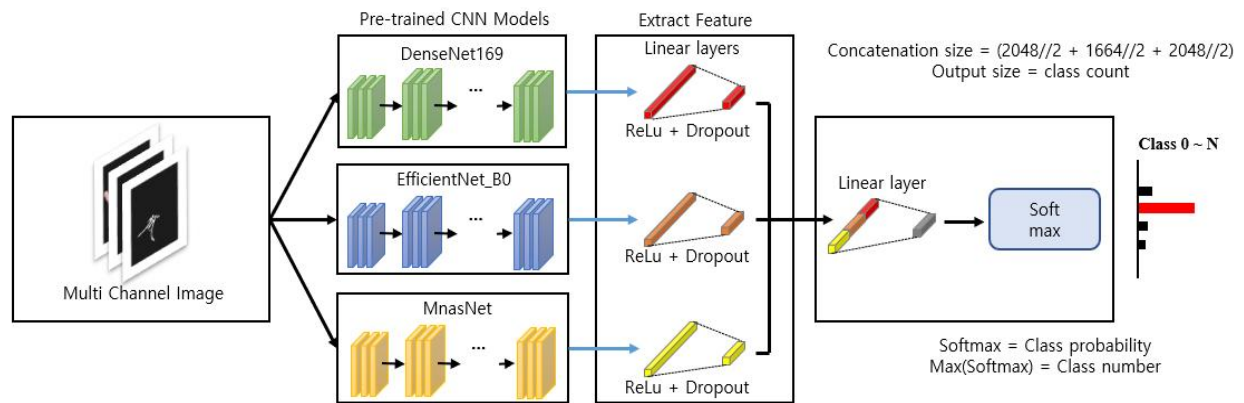


Fig. 12. Identification Model Architecture using Multi-channel images

사전 학습된 단일 CNN 모델들을 통해 Feature를 추출하고 Fully-Connected Layer에 학습한 결과 Best Top-3 모델은 DenseNet169, EfficientNet_B0, MnasNet 3개의 모델이 선정되었으며, Top-3 모델에 앙상블을 이용하여 Feature extraction을 진행하면 Epoch 30 기준 Best Loss Score는 0.2338로 단일 CNN 모델보다 앙상블 모델이 신원 인식에 더 적합하다. Fig. 13은 앙상블 모델에 대한 성능 지표를 보여준다.

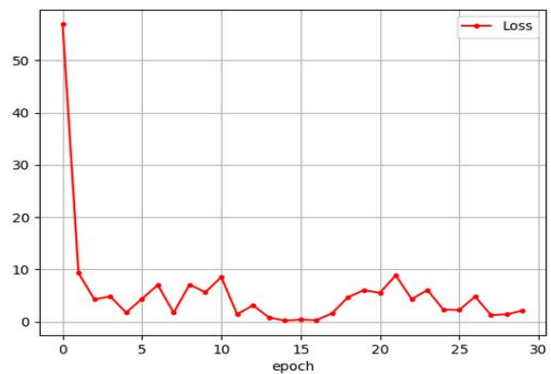


Fig. 13. Identification Model Metric

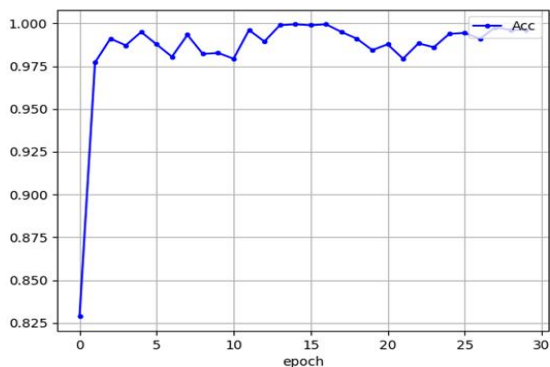


Fig. 14와 Table 7은 Test 데이터셋으로 A부터 G까지 7명의 신원 인식을 진행하였을 경우 나타나는 히트맵과 Precision, Recall, F1-Score, Accuracy이다.

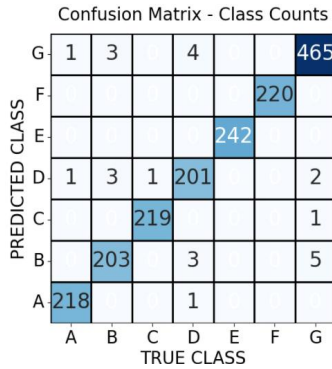


Fig. 14. Confusion Matrix of the Proposed Identification Model

Table 7. Identification Model Classification Performance

Class	Precision	Recall	F1-Score	Accuracy	# of images
Person A	99.54%	99.09%	99.32%	99.09%	220
Person B	96.21%	97.13%	96.67%	97.12%	207
Person C	99.55%	99.55%	99.55%	99.54%	219
Person D	96.63%	96.17%	96.40%	96.17%	207
Person E	100%	100%	100%	100%	244
Person F	100%	100%	100%	100%	220
Person G	98.31%	98.31%	98.31%	98.30%	478
Average	98.60%	98.61%	98.61%	98.61%	1793

이를 통해 본 연구에서 제안하는 멀티채널 이미지 기반 손바닥 신원 인식 모델은 7개의 클래스를 학습에 따라 평균 98.61%의 정확도로 예측하는 것을 확인할 수 있다. 클래스별로 생성된 Multi-channel image는 모두 20대 초, 중반 남성의 손바닥을 촬영하여 재구성한 이미지이다. 즉, 본 논문에서 제안하는 방법은 손바닥을 기반으로 정보를 추출하고 독립적인 특징들을 멀티채널 이미지로 재구성하였을 때 신원 인식에 효과가 있는 것을 보여준다.

IV. Conclusions

본 연구에서는 스마트폰 카메라 센서만을 이용하여 시스템 설치 비용, COVID-19 사태 등의 문제점을 해결하고 손바닥 정보를 이용하여 멀티채널 이미지를 생성하여 신원 인식에 사용 가능한 모델을 제안하며 멀티채널 이미지가 신원 인식 모델에 적합하다는 것을 보여주었다. 지금은 손가락을 포함한 손바닥, 손가락을 제외한 손바닥, 손금 3가지의 Segmentation 결과를 이용하여 멀티채널을 구성

하였지만, Landmark 등을 이용하여 길이, 구조 등의 정보를 추가하여 3채널 이상의 정보를 이용한다면 더욱 높은 보안성을 가질 것이다. 또한, GAN 등의 모델을 추가하여 제안된 아키텍처를 재구성한다면 더 복잡한 멀티채널 이미지에서도 좋은 성능이 나올 수 있다고 예상된다. 향후 연구 계획은 이처럼 손바닥 정보를 더욱 세분화하여 4채널 이상의 멀티채널 이미지를 활용하여 신원 인식 알고리즘의 성능을 고도화시킬 계획이다.

ACKNOWLEDGEMENT

This results was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-001(134 5341783)).

REFERENCES

- [1] J-M. Park, and B-J. Park, "The Password base System for the safe and Efficient Identification," Journal of the Korea Institute of Information and Communication Engineering Vol. 13, No. 1, pp. 81-86, January 2009, DOI: 10.6109/JKIICE.2009.13.1.81.
- [2] Y. Fu and W. Liu, "A New RFID-USB Key," 2007 International Workshop on Anti-Counterfeiting, Security and Identification (ASID), pp. 440-443, April 2007, DOI: 10.1109/IWASID.2007.373674
- [3] Y. Lee, "Security Enhancement to an Biometric Authentication Protocol for WSN Environment," Convergence Security Journal, Vol. 16, No. 6, pp. 83-88, 2021
- [4] H. Sung, and Y. Lee, "On-line Signature Identification Based on Writing Habit Information," Proceedings of the Korean Information Science Society Conference, Vol. 30, No. 1, pp. 322-324, 2003
- [5] M. M. H. Ali, V. H. Mahale, P. Yannawar and A. T. Gaikwad, "Overview of fingerprint recognition system," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 1334-1338, March 2016, DOI: 10.1109/ICEEOT.2016.7754900
- [6] H-S. Kim, and J-H. Cho, "A Method for Finger Vein Recognition using a New Matching Algorithm," Journal of KIISE, Vol.37, No.11, pp. 859-865, November 2010
- [7] G-W. Choi, and Y-J. Jeong, "Efficient iris recognition using deep-learning convolution neural network(CNN)," Journal of

- KIECS, Vol.15, No.3, pp. 521-526, 2020
- [8] S-E. Hong, W-B. Im, J-W. Park, and H-S. Yang, "Deep CNN-based Person identification using Facial and Clothing Features," Journal of IEIE, Vol.2016, No.6, pp. 2204-2207, June 2016
- [9] J. Gwak "Multi-object tracking through learning relational appearance features and motion patterns," Computer Vision and Image Understanding, Vol.162, pp. 103-115, September 2017, DOI: 10.1016/j.cviu.2017.05.010
- [10] E. Yang, J. Gwak, and M. Jeon "Conditional random field (CRF)-boosting: constructing a robust online hybrid boosting multiple object tracker facilitated by CRF learning," Sensors, Vol.17, No.3, March 2017, DOI: 10.3390/s17030617
- [11] Y-H. Choi, "Focus on the next generation of biometric security technologies," KIRS, <https://ssl.pstatic.net/imgstock/upload/research/industry/1627526012944.pdf>, July 2021
- [12] S-B. Kim, and W-J. Kim, "User Identification Method using Palm Creases and Veins based on Deep Learning," Journal of Broadcast engineering, Vol.23, No.3, pp. 395-402, May 2018
- [13] L. G. Oldal and A. Kovács, "Hand geometry and palmprint-based authentication using image processing," 2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY), pp. 125-130, 2020, DOI: 10.1109/SISY50555.2020.9217068.
- [14] G. Jaswal, A. Kaul, and R. Nath, "Multiple feature fusion for unconstrained palm print authentication," Computers & Electrical Engineering, Vol.72, pp. 53-78, September 2018, DOI: 10.1016/j.compeleceng.2018.09.006
- [15] T. P. Van, S. T. Nguyen, L. B. Doan, N. N. Tran, and T. M. Thanh, "Efficient Palm-Line Segmentation with U-Net Context Fusion Module," International Conference on Advanced Computing and Applications(ACOMP), pp. 23-28, November 2020, DOI: 10.1109/ACOMP50827.2020.00011
- [16] S. C. Soh, M. Z. Ibrahim, and M. Yakno, "A Review: Personal Identification Based on Palm Vein Infrared Pattern," Journal of Telecommunication, Electronic and Computer Engineering(JTEC), Vol.10, No.10, pp. 175-180, January 2018
- [17] O. Oktay, J. Schlemper, L. Le Folgoc, L. Matthew, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Hammerrla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net : Learning Where to Look for the Pancreas," 2018
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net : Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention, pp. 234-241, May 2015
- [19] H. Lim, and J. Gwak, "Generative optical flow based abnormal object detection method using a spatio-temporal translation network," Journal of the Korea Society of Computer and Information, Vol.26, No.4, pp. 1-8, April 2021, DOI: 10.9708/JKSCI.2021.26.04.011
- [20] J. Kang, and J. Gwak, "Adaptive Face Mask Detection System based on Scene Complexity Analysis," Journal of the Korea Society of Computer and Information, Vol.26, No.5, pp. 1-8, May 2021, DOI: 10.9708/JKSCI.2021.26.05.001
- [21] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," (2012) In Proceedings of the 25th International Conference on Neural Information Processing Systems, Vol.60, No.6, pp. 84-90, May, 2017, DOI: 10.1145/3065386
- [22] G. Huang, Z. Liu, and Q. Killan, "Densely Connected Convolutional Networks," In Proceedings of the IEEE conference on computer vision and pattern recognition p.4700-4708, 2017, DOI: 10.1109/cvpr.2017.243
- [23] M. Tan, and V. Quoc, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," In International conference on machine learning, pp. 6105-6114, May 2019
- [24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, 2016, DOI: 10.1109/CVPR.2016.308.
- [25] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, V. Quoc "MnasNet : Platform-Aware Neural Architecture Search for Mobile", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2820-2828, 2019
- [26] G. Andrew, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, April 2017
- [27] K. He, X. Zhang, S. Ren, and J. Sun "Deep Residual Learning for Image Recognition," In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016, DOI: 10.1109/CVPR.2016.90
- [28] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He "Aggregated Residual Transformations for Deep Neural Networks," In Proceedings of the IEEE conference on computer vision and pattern recognition pp. 1492-1500, 2017
- [29] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6848-6856, 2018
- [30] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, September 2014
- [31] J. Kang, and J. Gwak, "Deep Learning-Based Brain Tumor Classification in MRI Images Using Ensemble of Deep Features," Journal of the Korea Society of Computer and Information, Vol. 26, No. 7, pp. 37-44, July 2021, DOI: 10.9708/JKSCI.2021.26.07.037

Authors



Inki Kim received the B.S degree in Computer engineering from Korea National University of Transportation (KNUT) in 2020. From 2022, he has been working as a Master's student and researcher Artificial Machine Intelligence

Lab in KNUT. His research interest is deep learning, computer vision, and multimodal data segmentation using machine learning and deep learning.



Beomjun Kim received the B.S degree in Computer engineering from Korea National University of Transportation (KNUT) in 2021. From 2022, he has been working as a Master's student and researcher Artificial

Machine Intelligence Lab in KNUT. His research interest is deep learning, computer vision, and segmentation using machine learning and deep learning.



Sunghee Woo received the Ph.D. degree in Electrical engineering from Chungbuk University, Cheongju, Korea in 1999. Dr. Woo joined Korea National University of Transportation in 1995 as an assistant professor. She is currently a Processor.

Her's current research interests include information security, computer network, and big data processing/analysis.



Jeonghwan Gwak received the Ph.D. degree in machine learning and artificial intelligence from Gwangju Institute of Science and Technology (GIST), Gwangju, Korea in 2014. From 2002 to 2007, he had worked for several companies and

research institutes as a researcher and a chief technician. From 2014 to 2016, he worked as a postdoctoral researcher in GIST, and from 2016 to 2017 as a research professor. From 2017 to 2019, he was a research professor in Biomedical Research Institute & Department of Radiology at Seoul National University Hospital, Seoul, Korea. From 2019, he joined Korea National University of Transportation as an assistant professor, and he is the director of the Applied Machine Intelligence laboratory. His current research interests include deep learning, computer vision, signal and image processing, AIoT, evolutionary algorithms and optimization, fuzzy sets and systems, and relevant applications of biomedical and visual surveillance systems.