Journal of KIIT. Vol. 22, No. 6, pp. 41-53, Jun. 30, 2024. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2024.22.6.41

이미지 특징 강화와 앙상블 학습을 활용한 디지털 조작 이미지 검출 프레임워크

길현선*[1], 김지호*[2], 이홍철**

Digital Image Forgery Detection Framework based on Image Feature Enhancement and Ensemble Learning

Hyunsun Kil*[1], Jiho Kim*[2], and Hongchul Lee**

본 연구는 4단계 두뇌한국21에 의해 지원되었습니다.

요 약

디지털 환경에서 조작 이미지는 무분별하게 생성되며, 이에 따라 적응성과 범용성을 갖춘 자동화된 검출 시스템이 요구된다. 기존의 연구에서는 특정 형태의 조작만을 검출하거나 상황에 국한된 모델을 제안하여 다양한 환경에 확장이 어렵다는 한계가 존재하였다. 또한 최신 조작 경향성을 반영한 데이터를 활용한 연구가 부족한 상황이다. 따라서 본 연구는 다양하고, 새로운 조작 이미지를 효과적으로 검출하고, 확장 가능성을 갖춘 강건한 딥러닝 프레임워크를 제안한다. 이를 위해 조작된 이미지의 특징을 강화하였으며 다수의 딥러닝 기반 모델에 앙상블 학습을 적용하였다. 주요 결과로는 컨볼루션 신경망(Convolutional Neural Network)과 비전 트랜스포머(Vision Transformer) 기반의 다섯 개의 모델을 앙상블 한 결과 기본 모델과 비교하여 최소 3.45%에서 최대 10.4%의 정확도가 상승하였으며, 특징을 강화한 조작 이미지가 검출 성능에 효과적임을 확인하였다.

Abstract

In the digital environment, forged images are generated indiscriminately, requiring automated detection system with adaptability and versatility. The limitations of existing studies are that they detect only certain types of manipulation or propose models that are limited to a specific situation, making it difficult to extend to various environments. In addition, there is a lack of research utilizing data that reflects the latest forgery trends. Therefore, This paper proposes a scalable and robust deep learning-based framework for effectively detecting diverse and novel types of digitally forged images. To address this we applied feature enhancing on forged images and ensemble learning on multiple deep learning-based models. The main results are that the ensemble of five Convolutional Neural Network and Vision Transformer-based models showed an accuracy increase of at least 3.45% and up to 10.4% compared to the baseline model, and confirmed that feature-enhanced image was effective in improving detection performance.

Keywords

computer vision, image forgery detection, ensemble learning, feature enhancement, CNN, vision transformer

* 고려대학교 산업경영공학과
- ORCID[1]: https://orcid.org/0009-0004-4414-6738
- ORCID[2]: https://orcid.org/0000-0003-3733-8702
** 고려대학교 산업경영공학과 교수(교신저자)
- ORCID: https://orcid.org/0000-0002-4407-0348

· Received: Apr. 03, 2024, Revised: May 16, 2024, Accepted: May 19, 2024
· Corresponding Author: Hongchul Lee
Dept. of Industrial and Management Engineering, Korea University, Korea
Tel.: +82-2-3290-3389, Email: hclee@korea.ac.kr

Ⅰ. 서 론

디지털 기술의 발전으로 컴퓨터와 스마트폰 등을 통해 생성되고 공유되는 이미지의 양이 기하급수적으로 증가하고 있다. 누구나 쉽게 이미지를 편집할 수 있게 됨에 따라 거짓 정보가 생성되고, 이로 인해 개인의 저작권과 초상권이 침해되는 것을 넘어 더 큰 피해가 발생하고 있다. 특히 이미지 조작에 의한 피해는 발생 이전으로 돌이킬 수 없다는 점에서 해결이 필수적인 사회문제로 여겨진다[1]. 인스타그램(Instagram)과 같은 소셜 네트워크 서비스(SNS, Social Network Service)에서만 하루 950만 개 이상의 시각 콘텐츠가 공유되고, 2020년에는 약 1.4조 개의 디지털 사진이 생산된 것으로 추정된다[2]. 끊임없이 쏟아져 나오는 조작된 디지털 이미지를 사람이 일일이 검증하기에는 명백한 한계가 존재하며, 이에 따라 조작 여부를 빠르게 판별할 수 있는 이미지 필터링 시스템이 필요한 실정이다.

조작된 디지털 이미지는 원본 이미지가 허위 혹은 부정하게 변경된 이미지를 말한다. 이미지 조작은 주로 인페인팅(Inpainting), 복사·붙여넣기(Copy-paste), 접합(Splicing), 보정(Retouching)의 네 가지 방법으로 이루어진다. 인페인팅은 이미지의 일정 부분을 이웃하는 픽셀들과 일관되게 수정하는 조작 방식이며, 복사·붙여넣기는 하나의 이미지 내에서 일부를 옮겨서 붙여 넣는 조작 방식이다. 접합은 두 개 이상의 각각 다른 이미지를 하나의 이미지 내에 붙이는 방식이고, 보정은 원본 이미지에 각종 편집 효과를 적용하는 방법이다[3]. B. Shah et al.[4]와 K. R. Revi et al.[5]에서는 복사·붙여넣기와 접합을 대표적인 이미지 조작 방식으로 제시하며, 특히 K. R. Revi et al.[5]에서는 인물 사진에서의 접합을 검출하여 개인의 명예를 실추시키는 문제를 예방하였다. K. B. Meena et al.[6]은 조작의 형태를 복사·붙여넣기, 접합, 리샘플링(Resampling), 보정의 네 가지로 분류하며, 보정은 불법적인 목적으로 사용되는 경우가 드물다고 주장하였다.

이미지 조작 검출에 관한 연구는 Image forgery detection, Image tampering detection, Image manipulation detection, 디지털 포렌식 등의 분야에서 활발히 진행되고 있다. 각 분야에서는 주로 두 가지 접근 방식을 사용하여 이미지 훼손을 확인하며, 이는 이미지에 특정한 증명을 심어 원본 여부를 확인하는 능동적 접근 방식과 이미지에서 추출된 시각적 특징을 통해 이상 여부를 확인하는 수동적 접근 방식이 존재한다[7]. 능동적 접근 방식은 디지털 포렌식 분야에서 디지털 콘텐츠의 일관성을 유지하는 연구에서 활용되며, 수동적 접근 방식은 이미지 분류, 영상 분할 등 목적에 따라 다양한 분야의 전통적인 방법론을 비롯하여 머신러닝, 딥러닝 기반의 컴퓨터 비전 연구가 활용되고 있다[3]. 조작 검출 분야 또한 조작 여부를 판별하거나, 조작 위치를 찾아내는 두 가지 연구 방향으로 나뉘어 수행된다[8].

딥러닝은 인공지능 분야에서 가장 활발히 연구되고 있는 분야로, 이미지 조작 검출에서도 뛰어난 성과를 보인다[9]. 딥러닝 기반의 이미지 처리 기술은 데이터의 고유한 구조적 패턴에 맞춰 자동으로 특징을 추출하여, 기존의 방법과 비교했을 때 더욱 강력한 성능을 보인다[8]. 최근 이미지 조작 검출 문제에 딥러닝을 적용한 연구가 활발히 이루어지고 있으며, 특히 컨볼루션 신경망(CNN, Convolutional Neural Network) 기반 단일 모델[10]-[13]을 활용하거나 특정 조작 형태에 관한 검출 연구가 다수 수행되었다[14][15].

본 논문은 조작 이미지 검출을 위한 효과적인 딥러닝 앙상블 프레임워크를 제안하였다. 조작 이미지의 에지 특징을 강화하고 다양한 딥러닝 알고리즘의 분류 결과를 결합하여 다양성과 강건함이 요구되는 이미지 조작 검출 분야에 효과적이고 확장이 쉬운 시스템을 제안하였다.

논문의 구성은 다음과 같다. 2장에서는 딥러닝 기반 조작 이미지 탐지 및 앙상블 학습 관련 연구에 관해 설명하고, 3장에서는 본 연구에서 사용한 데이터 세트, 특징 강화 기법, 딥러닝 방법론 및 앙상블 기법을 소개한다. 4장에서는 실험 환경, 평가 지표 및 결과를 다루며, 5장에서는 결론과 향후 연구에 관해 기술한다.

Ⅱ. 관련 연구

2.1 Deep learning-based image forgery detection

최근 딥러닝은 다양한 분야에서 좋은 성능을 보이며, 5년간 딥러닝 기반 방식이 전통 방식의 이미지 조작 검출 성능을 뛰어넘고 있다[8]. 이미지 분류에 가장 널리 쓰이는 모델은 CNN 기반 모델과 Vision Transformer 기반 모델로 VGGNet[16], ResNet[17], EfficientNet[18], Vision Transformer (ViT)[19] 등이 주요 모델이나 특징 추출에 사용되며, 조작 이미지 탐지에도 활용되고 있다.

N. H. Rajini[20]은 CASIA v1.0, v2.0 데이터를 활용하였고, 제르니케 모멘트(Zernike moment polar)와 구획 이산 코사인 변환(Block discrete cosine transform)을 사용하여 CNN 기반의 모델의 특징 추출에 기여하였다. Abhishek et al.[21]은 CNN 기반의 모델을 2개의 단계로 배치하여 분류 성능을 향상했다. S. Kumar et al.[22]는 VGG와 Inception-v3 모델을 결합한 하이브리드 모델로 특징을 결합하여 분류를 수행하였는데 4개의 각각 다른 데이터 세트를 활용하여 모델이 학습되었으며 CoMoFoD[23] 데이터에 검증 및 테스트하여 99%의 정확도를 도출하였다. M. Aria et al.[24]는 컨볼루션 계층을 쌓아서 구성하는 GAN을 활용하여 조작 여부를 탐지하고, 조작된 위치를 탐지하는 연구를 수행하였다. S. Ganguly et al.[25]는 Vision Transformer와 Xception Network를 활용하여 딥페이크(DeepFake) 조작에 높은 성능을 보였다.

2.2 Applications of Ensemble Learning

앙상블 학습은 다중 결정자를 사용해 단일 모델의 한계를 극복하고, 과적합을 방지하여 예측력을 높이는 학습 방식으로, 의료, 음성 인식, 이미지 분류, 시계열 예측 등 다양한 분야에 적용되며, 다수의 연구를 통해 효과가 증명되었다.

H. Kashiani et al.[26]은 ViT 와 EfficientNet, ResNet 모델을 다양한 방식으로 앙상블 하여 Morph attack 검출 성능을 향상하였고, A. Hashmi et al.[27]은 오디오와 비디오 데이터를 활용한 멀티모달(Multi modal) 조작 검출의 성능 향상에 보팅 앙상블을 사용하였다. K. R. Revi et al.[5]는 VGG, ResNet을 다수결(Majority voting) 앙상블 하여 얼굴 사진에 대한 접합 검출에 좋은 성능을 보였고, E. Tasci et al.[28]은 ResNet과 VGG를 포함한 3개의 모델을 소프트 보팅(Soft voting) 및 베이지안 최적화 기반 가중 투표(Bayesian optimization-based weighted voting) 앙상블 하여 결핵 탐지 성능을 향상하였다.

그러나 여전히 딥러닝 모델을 사용하는 조작 이미지 검출연구가 많지 않으며, 특히 다중 모델을 활용한 연구의 수가 적은 상황이다. 따라서 본 연구는 기존 연구에서 효과적인 성능을 보인 딥러닝 모델과 최신 디지털 조작 이미지에 특징 강화 기법을 사용하여 다양성과 강건함을 제고할 수 있는 딥러닝 기반의 조작 이미지 앙상블 프레임워크를 제안한다.

III. 디지털 이미지 조작 검출 프레임워크

3.1 Overall process

본 연구에서는 SNS상의 실제 조작 형태를 반영한 이미지 조작 검출 프레임워크를 제안하며 전체 구조는 그림 1과 같다. 제안하는 프레임워크는 총 2개의 모듈로 구성된다. 첫 번째는 특징 강화 모듈로 이미지 내 부자연스러운 조작 경계를 강조하여 효율적으로 포착할 수 있게 한다. 두 번째는 이미지 조작 검출 모듈로 가장 높은 성능의 5개의 이미지 분류 모델을 앙상블 하여 이미지의 조작 여부를 판별한다.

3.2 Dataset

본 연구에서는 세계 최대의 소셜 미디어 회사인 Meta가 2021년 공개한 Image Similarity Challenge 2021 dataset (DISC21)[29]를 사용한다. 원본과 조작 이미지를 포함하여 2,100,000개의 이미지로 구성되어 있다. DISC21은 YFCC100M[30], DeepFake Detection Challenge dataset[31], Casual Conversations Dataset[32]에서 유래되어 복사·붙여넣기, 접합, 보정의 방식으로 조작되었으며, 구체적으로 오버레이(문자, 이모티콘) 삽입, 색상(밝기, 채도, 흑백, 필터 적용 등) 변경, 픽셀 단위 조작(흐림, 디더링(Dithering), JPEG 인코딩 등), 공간적 변형(자르기, 회전, 여백 삽입 등) 등의 방식으로 조작되었다. 소셜 미디어상의 다양한 실제 사례가 반영된 데이터 세트의 예시는 그림 2와 같다.

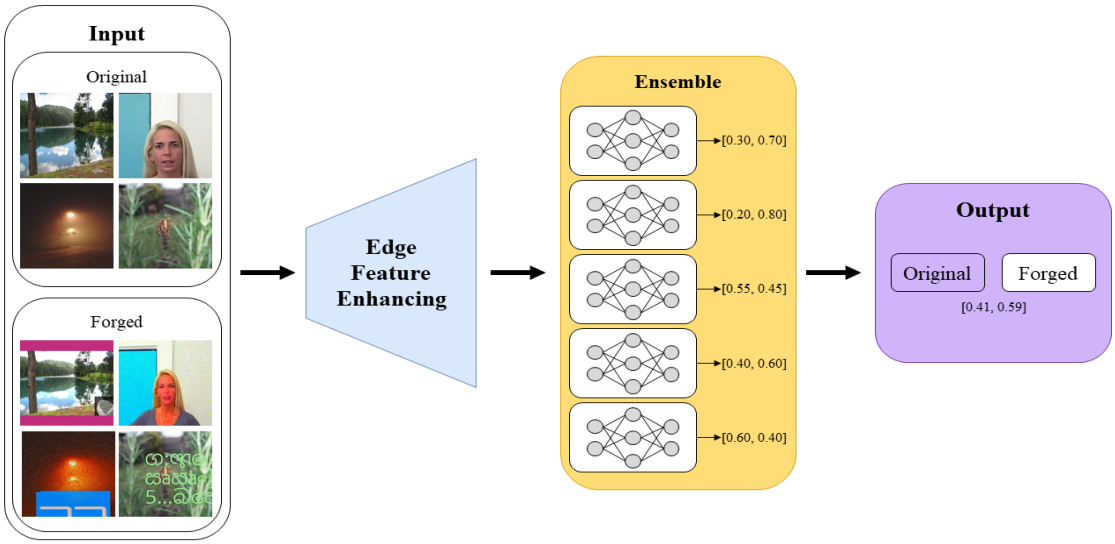

그림 1. 전체적인 구조도
Fig. 1. Overall architecture

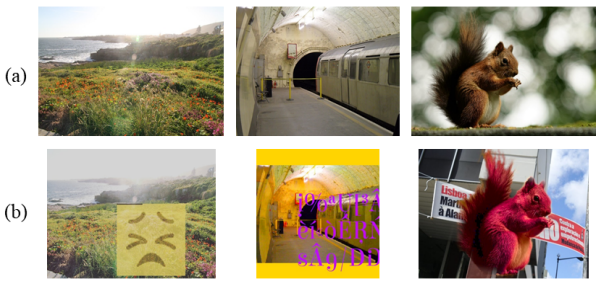

그림 2. 원본 이미지와 조작 이미지 예시
(a) 원본 이미지, (b) 조작 이미지
Fig. 2. Example of original and forged image
(a) Original image, (b) Forged image

실험을 위해 표 1과 같이 데이터 세트를 구축하였다. 편향된 결과 도출을 방지하기 위해 무작위로 추출한 20,000장으로 데이터 세트를 구축하였으며, 8:1:1의 비율로 나누어 실험을 진행하였다.

표 1. 학습 및 실험 데이터 세트
Table 1. Summary of the train and test dataset

DISC21 dataset		Train	Validation	Test
Class	Images			
Original	10,000	8,000	1,000	1,000
Forged	10,000	8,000	1,000	1,000
Total	20,000	16,000	2,000	2,000

3.3 Feature enhancements

이미지 편집 기술은 꾸준히 발전하여 인공지능을 활용한 이미지 편집은 실제 사진과 차이를 느끼기 어려울 수준으로 발전이 이루어지고 있다. 그러나 여전히 조작된 이미지 대부분은 조작의 경계가 자연스럽지 않기 때문에 조작 경계를 포착하는 방법으로 조작 검출이 가능하다. 이미지 내에 존재하는 물체 간 경계를 뚜렷하게 만드는 알고리즘은 다양하게 존재하며, 본 연구에서는 적응적 임계화(Adaptive thresholding)와 캐니 에지 검출(Canny edge detection) 알고리즘을 사용하여 이미지 내 에지 특징을 강화하여 효과적인 이미지 조작 검출을 수행하고자 한다.

그림 3에서 원본 이미지 (a), 조작 이미지 (b), 적응적 임계화와 캐니 에지 검출 알고리즘을 각각 사용하여 추출된 에지 특징 이미지 (c), (e)와 에지 특징이 강화된 이미지 (d), (f)를 확인할 수 있다.

[33]에서는 이미지의 임곗값에 따라 영상을 어둡게 또는 밝게 처리하는 방법을 제안한다. 그러나 이미지 전체에 하나의 임곗값을 적용하게 되면 조도에 따른 이미지 내의 부분적 차이를 감지하지 못하기 때문에 영역별로 다른 임곗값을 설정하여 조명의 변화나 반사가 심한 경우에 효과적인 적응적 임계화를 활용하였다. A. S. Methil et al.[34]는 적응적 임계화를 사용하여 자기 공명 영상(MRI) 이미지의 에지를 추출하여 뇌종양 검출 성능을 높였으며, V. Ribeiro et al.[35]는 차량 번호판 검출 성능을 향상했다.

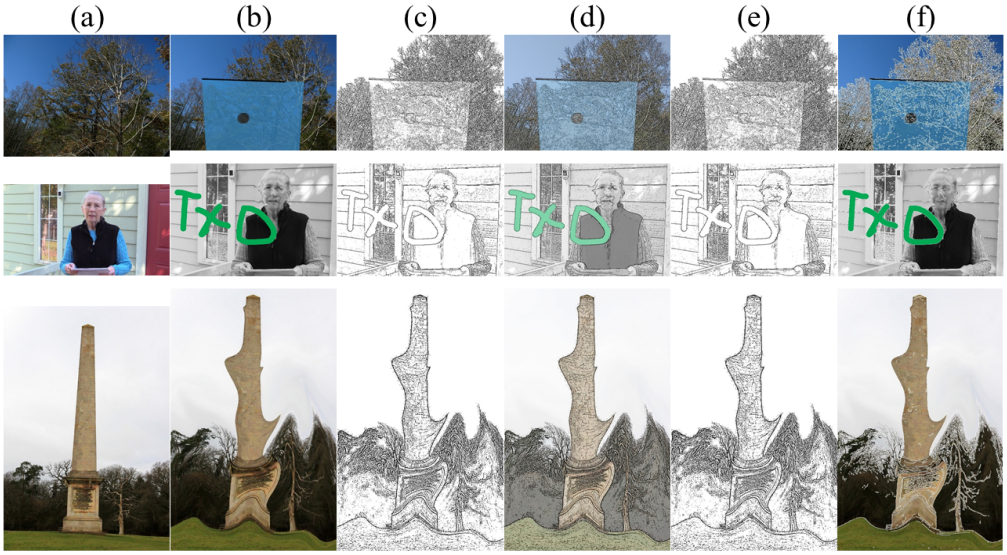

그림 3. 추출된 특징과 특징이 강화된 이미지
Fig. 3. Examples of feature extracted and enhanced images

$I(x,y)$가 흑백 입력 이미지에 대한 픽셀의 밝기일 때, 평균을 사용하는 적응적 임계화의 이진 임곗값 $T(x,y)$는 식 (1)과 같으며 이웃한 픽셀의 값과 계산되고, 이웃은 블록의 크기에 의해 범위가 결정된다. 임곗값에 따라 나누어진 흑백의 출력 이미지 $dst(x,y)$는 식 (2)와 같으며 이러한 연산 과정은 이미지 내의 각 픽셀에 적용되어 결과적으로 배경과 전경이 분리된 이미지를 얻는다.

$$T(x,y) = mean(I(x,y)) - C \tag{1}$$

$$dst(x,y) = \begin{cases} max\,Value, & \text{if } I(x,y) > T(x,y) \\ 0, & otherwise \end{cases} \tag{2}$$

캐니 에지 검출[36]은 이미지의 픽셀마다 기울기(Gradient)를 계산하여 수직, 수평 및 대각선 에지를 찾아내고, 최소와 최대 임곗값에 맞는 에지를 남기는 연산을 수행한다. 에지 검출에 가장 널리 사용되며, 얇고, 정확도 높은 에지를 추출하는데 효과적인 캐니 에지 검출 알고리즘을 활용하였다.

캐니 에지 검출은 식 (3)과 같이 픽셀의 좌표 x, y 에 대해 가우시안(Gaussian) 필터를 통해 이미지에 존재하는 잡음을 제거하고, 식 (4), (5)에서와 같이 수직과 수평의 기울기 G_x, G_y를 활용해 에지의 기울기와 각도를 계산한다.

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^{2+}y^2}{2\sigma^2}\right) \tag{3}$$

$$Edge\ Gradient\ (G) = \sqrt{G_x^2 + G_y^2} \tag{4}$$

$$Gradient\ Angle(\theta) = \tan^{-1}\frac{G_y}{G_x} \tag{5}$$

이후 기울기 벡터와 같은 방향의 인접 픽셀 중 가장 변화율이 높은 픽셀만 에지로 남기고, 임곗값과 비교하여 최종 에지로 판단하거나, 추가적인 검사를 수행하여 얻어진 에지를 연결해가며 강조된 에지 이미지를 얻는다.

3.4 Deep learning-based image forgery detection models

본 연구에서 사용된 CNN 기반 모델은 총 7개로, VGG16[16], VGG19[16], ResNet-50[17], ResNet-152[17], MobileNetV2[37], EfficientNet-b0[18], EfficientNet-b7[18]이며, Transformer 기반 모델은 총 2개로, ViT[19], Swin Transformer[38]이다. CNN과 Vision Transformer의 기본 구조는 그림 4, 그림 5와 같으며, 신경망 기반 모델은 같은 이미지에 대해서도 모델마다 각자 다른 형태의 특징을 추출하며, 학습 효율과 성능을 개선하기 위한 다양한 방법론이 제안되었다.

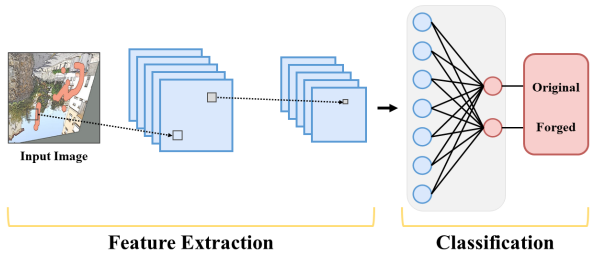

그림 4. CNN의 구조
Fig. 4. Architecture of CNN

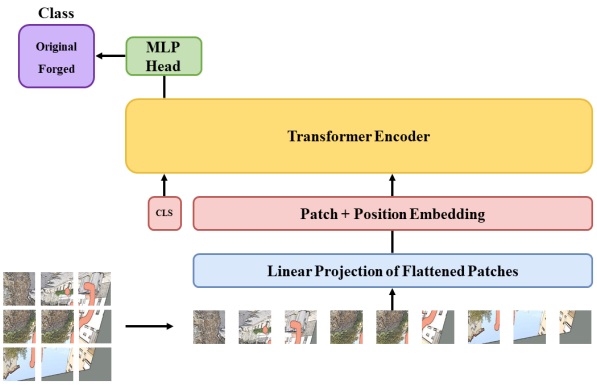

그림 5. Vision Transformer의 구조
Fig. 5. Architecture of Vision Transformer

각각의 모델은 조작 이미지 탐지 분야에 꾸준히 활용되고, 최신의 선행 연구에서 검증된 분류 성능[39][40]을 고려하여 선택되었으며, 같은 모델 내에서도 얕은 구조와 깊은 구조를 모두 활용하여 정확도를 극대화하였다. 각 모델의 특징은 다음과 같다.

VGGNet은 대용량 벤치마크 데이터 세트인 ImageNet에 이미지 분류로 SOTA(State-of-the-Art) 성능을 낸 모델로 네트워크 전체에 3×3 컨볼루션 필터를 적용하고, 컨볼루션 계층을 증가시켜 복잡한 패턴을 포착할 수 있다. ResNet은 신경망의 계층 사이에 Skip connection을 통해 기울기 소실 문제를 해결하여 빠른 학습과 높은 일반화 성능을 가지며, MobileNetV2는 Inverted residual block을 통해 에지 디바이스에 적합한 모델로 계산 효율이 높고, 적은 메모리로 빠른 학습이 가능하다. EfficientNet은 Compound scaling으로 기존의 CNN 기반 방식에서 성능 향상을 위해 깊은 네트워크, 채널 증가, 입력 해상도 증가 등 다양한 스케일업(Scale-up) 방식에 AutoML을 적용하여 최적의 조합으로 높은 성능과 효율을 가진다. ViT는 컴퓨터 비전 분야에 Transformer 구조를 사용하여 CNN의 지역 패턴 학습과 달리 이미지 전체의 맥락을 학습할 수 있다. Swin Transformer는 임베딩(Embedding) 과정에서 패치를 합쳐가며 계층적 구조를 구성하며, 윈도(window) 내에 포함된 패치 간의 Self-attention 후 윈도를 이동하는 방식으로 순차적으로 연산하여 장기 정보 손실을 해결하고, 확장성과 대규모 데이터 처리가 가능하다. 각 모델의 복잡도는 표 2와 같다.

표 2. 모델 복잡도
Table 2. Complexity of models

Models	Parameters	Layers
VGG16	138M	16
VGG19	144M	19
ResNet-50	2.5M	50
ResNet-152	6M	152
MobileNetV2	3.4M	53
EfficientNet-b0	5.3M	237
EfficientNet-b7	6.6M	813
ViT	86.6M	12
Swin transformer	87.8M	24

3.5 Ensemble learning

본 연구에서는 효과적인 이미지 조작 검출 시스템을 위한 앙상블 프레임워크를 제안한다. 이미지 조작 방식은 다양하게 존재하기 때문에 단일 모델에서 도출된 결과보다 다중 모델의 결과를 결합하여 분류 정확도와 탐지 다양성을 확보할 수 있는 앙상블 프레임워크를 구성하였다.

앙상블에 활용되는 알고리즘과 모델의 구성은 문제에 의존적이므로 연구자에 따라 기준이 달라지는 특징을 가진다[41]. 본 연구에서는 전체 앙상블 기법을 적용한 연구 중 30%가량 활용된 결정 융합(Decision fusion)의 일종인 소프트 보팅을 통해 5개 모델의 결과를 독립적으로 결합하여 확장성을 갖춘 앙상블 모델로 평균 예측 확률을 도출하여 결과 레이블을 출력하였다. 3개, 4개, 5개의 모델을 결합한 성능을 확인한 결과, 3개 혹은 5개를 결합하였을 때 성능이 높게 나왔으며, 모델의 장점을 활용해 다양한 조작을 검출해내는 것이 목표이므로 5개의 모델을 결합하였다.

앙상블 모델의 연산은 다음과 같이 이루어진다. M은 앙상블에 사용된 모델의 개수이며, 데이터 x에 대해 $p_i^l(x)$는 모델 i에서 클래스 l을 예측할 확률이다. x가 l클래스에 속할 확률 $P_{ensemble}^l(x)$는 식 (6)과 같다. 각 분류기는 목표 클래스일 확률을 제공하고, 각 분류기에 할당된 가중치 α_i가 곱해져 각 분류기의 중요도를 반영한다. 본 연구에서는 가중치를 모두 1로 고정하였다. 최종적으로 식 (7)에서와 같이 가장 높은 확률을 가진 클래스를 최종 결과물로 산출한다.

$$P_{ensemble}^l(x) = \frac{1}{M}\sum_{i=1}^{M}\alpha_i p_i^l(x) \tag{6}$$

$$Predicted\,class = argmax_l P_{ensemble} \tag{7}$$

그림 6에서 본 연구에서 최종적으로 제안된 앙상블 모델의 구조를 나타낸다.

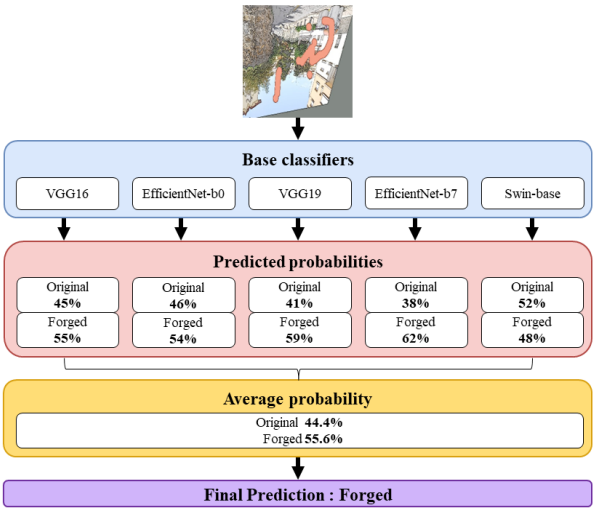

그림 6. 소프트 보팅 앙상블
Fig. 6. Soft voting ensemble

IV. Experiments

4.1 Experiment setting

4.1.1 Hyperparameter setting

실험을 위한 손실함수로 이진 및 다중 분류 문제에서 사용되는 교차 엔트로피(Cross-entropy)를 활용하였으며 이는 실제와 예측 간의 확률 차이를 통해 높은 확률로 예측하면 손실이 감소되게 하는 방식으로 예측 정확도를 향상시킨다. 옵티마이저(Optimizer)로 SGD에 momentum을 더하여 변동성을 줄이고, 지역 최적에 수렴하는 것을 방지하기 위해 조작 이미지 분류에서 사용되는 SGDM[48]을 활용하였다. 본 연구의 최적 하이퍼 파라미터(표 3)는 CNN과 Transformer 기반 모델의 학습과 테스트의 각 단계를 통해 결정되었으며 연구에서 사용된 모든 모델에 동일하게 적용되었다.

표 3. 최적의 하이퍼 파라미터
Table 3. Optimal hyperparameters

Input size	224×224×3
Optimizer	SGD
Learning rate	1e-3, 1e-2
Weight decay	1e-4
Optimizer momentum	0.9
Batch size	32, 64
Training epoch	100

4.1.2 Evaluation

본 연구의 목적은 이미지 조작 검출로 검출 여부를 분류하는 것으로 볼 수 있다. 따라서 평가 지표로는 식 (8)~(11)의 수식으로 표현된 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score를 사용하였다. 분류 결과 분석을 위한 혼동행렬은 표 4와 같다.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{8}$$

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$F_1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{11}$$

표 4. 혼동 행렬
Table 4. Confusion matrix

Confusion matrix		Predicted class	
		Original	Forged
True class	Original	TP	FN
	Forged	FP	TN

4.2 Experiment results

4.2.1 Feature enhancement results

본 연구에서는 특징 강화를 위해 적응적 임계화 및 캐니 에지 검출 알고리즘을 활용하였으며, 두 알고리즘을 함께 적용한 방법을 제안하였다. 조작 검출 성능 확인을 위해서는 제안한 CNN 및 Transformer 기반 9개의 모델을 활용하였다[16]-[19][37][38]. 제안한 3가지의 방법론을 적용하여 강화된 이미지 결과는 그림 7과 같다.

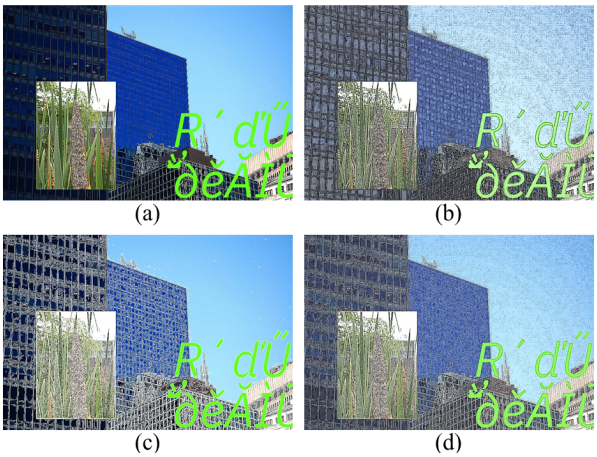

그림 7. 원본 이미지와 여러 에지 강화를 적용한 예시
(a) 조작 이미지, (b) 적응적 임계화,
(c) 캐니 에지 검출, (d) 모두 적용
Fig. 7. Example of original and feature enhanced image (a) Forged image, (b) Feature enhanced using adaptive thresholding, (c) Feature enhanced using canny edge detection, (d) Feature enhanced using both

표 5에서 9개의 모델에서 특징 강화를 통한 이미지 조작 검출 성능을 확인할 수 있다. 적응적 임계화를 통한 특징 강화 결과 9개 중 7개의 모델에서 최소 0.2%, 최대 9.15%, 평균 3.56%의 정확도가 향상되었다. 캐니 에지 검출을 통한 특징 강화 결과 9개 중 8개의 모델에서 최소 1.25%, 최대 6.45%, 평균 3.08%의 정확도가 향상되었다. 두 방식을 모두 적용한 경우 9개 중 3개의 모델에서 최소 0.35%, 최대 7.05%, 평균 2.82%의 정확도가 향상되었다. 단일 특징 강화를 적용한 것이 여러 가지를 동시에 적용한 것에 비해 높은 성능을 보이며, 적응적 임계화를 사용한 특징 강화가 이미지 조작 검출 성능 향상에 효과적임을 확인하였다. 이렇게 검증된 특징 강화를 다음 단계인 앙상블 모델에 적용하였다.

표 5. 특징 강화 기법 적용 결과 비교
Table 5. Comparison between experiment results of original and feature enhanced forged image

Models	Original	Adaptive thresholding	Canny edge detection	Using both
VGG16	78.90	**83.40**	80.80	79.95
VGG19	74.30	**83.45**	80.75	81.35
ResNet-50	75.70	**79.20**	78.75	75.30
ResNet-152	72.80	**77.70**	76.65	72.80
MobileNetV2	79.20	**79.40**	78.95	77.30
EfficientNet-b0	80.30	**81.65**	82.90	79.95
EfficientNet-b7	81.25	80.00	**82.70**	81.60
ViT	72.80	71.90	**76.85**	72.30
Swin transformer	78.25	**79.60**	79.50	72.90

4.2.2 Ensemble results

이미지 조작 검출 성능이 가장 우수한 모델 5개를 선택하여 앙상블을 수행하였으며 표 6은 제안된 최종 모델의 성능을 나타낸다.

표 6. 제안한 모델의 실험 결과
Table 6. Experiment results of proposed model

Models	Baseline				Feature enhanced using adaptive thresholding			
	Acc(%)	Prec(%)	Recall(%)	F1(%)	Acc(%)	Prec(%)	Recall(%)	F1(%)
VGG16	78.90	82.82	77.18	79.90	83.40	90.82	79.38	84.71
VGG19	74.30	76.21	73.88	75.02	83.45	86.77	81.69	84.16
EfficientNet-b0	80.30	88.35	76.43	81.96	81.65	86.38	79.26	82.66
EfficientNet-b7	81.25	83.71	80.15	81.89	80.00	88.06	76.17	81.68
Swin-base	78.25	78.87	78.33	78.60	79.60	83.02	78.09	80.48
Ours	-	-	-	-	**84.70**	89.54	**81.93**	**85.57**

Baseline 모델과 비교하여 적응적 임계화를 사용했을 때 VGG16 4.5%, VGG19 9.15%, EfficientNet-b0 1.35%, Swin-base 1.35%의 정확도가 향상되었다. 적응적 임계화를 사용한 후 앙상블 기법을 적용하였을 때 모든 기본 모델과 비교하여 최종적으로 정확도, 재현율, F1-score가 각각 84.70%, 81.93%, 85.57%로 향상되었다.

그림 8은 최종 결과를 시각적으로 표현한다. 주황색으로 표현된 제안 모델이 모든 평가 지표에 대해 기본 모델보다 최소 3.45%, 최대 10.4%의 월등한 성능 향상을 보였다. 실험결과는 특징 강화와 앙상블 학습이 이미지 조작 검출 성능을 향상했음을 보여준다.

표 7. 비교 실험 결과
Table 7. Comparative results

	Ours	[43]	[44]
Accuracy(%)	**84.70**	81	72.90
Base model	Ensemble	CNN	CNN
Dataset	DISC21	Dresden image database	CASIA-2.0, NC2016
Multi model	O	X	X
Training data	16,000	<14,000	13,000
Feature enhancement	O	X	X

표 7은 이미지 조작 검출을 수행한 선행연구와의 성능 및 특징을 비교한 것으로 본 연구에서 제안한 모델이 가장 효과적임을 확인하였다. [43], [44]에서는 다중 모델이 아닌 단일 CNN 기반의 모델을 활용하였고, 특징 강화를 적용하지 않았다. 따라서 본 연구에서 사용된 다중 모델 앙상블과 조작 이미지의 에지 특징 강화 방법론이 조작 이미지 탐지 성능 개선에 효과적이었음을 증명하였다.

또한 [10], [11], [13]에서는 단일 CNN 모델을 활용하였고, [20]에서는 단일 CNN 모델에 특징 강화를 적용하여 성능을 향상시켰다. [5]에서는 illuminant maps를 입력으로 VGG, ResNet을 앙상블하여 얼굴 사진에 대한 접합 검출 성능을 향상하였으며, [35]에서는 CLAHE를 통한 전처리와 ResNet과 VGG를 포함한 3개의 모델을 앙상블 하여 의료 분야에서 결핵 탐지 성능을 향상하였다.

본 연구의 프로세스는 기존의 단일 모델에 전처리를 추가하거나, 다중 모델에 앙상블만을 적용한 국지적인 수정이 아닌 전체 단계에 걸쳐 성능 향상과 다양한 모델을 활용한 확장성에 기여하는 연구로서 독창성을 가진다.

또한 본 연구에서 사용된 데이터는 SNS 상의 실제 조작을 반영하는 최신 데이터 세트인 것에 반해 기존의 조작 이미지 탐지 연구에 사용되는 데이터 세트는 CASIA 1.0, CASIA 2.0, CoMoFoD, MICC-F220, MICC-F600, MICC-F2000, Columbia 등으로 이는 10년 이상 된 조작 이미지 데이터로서 최신의 조작 트렌드 반영이 부족하다는 한계가 존재한다.

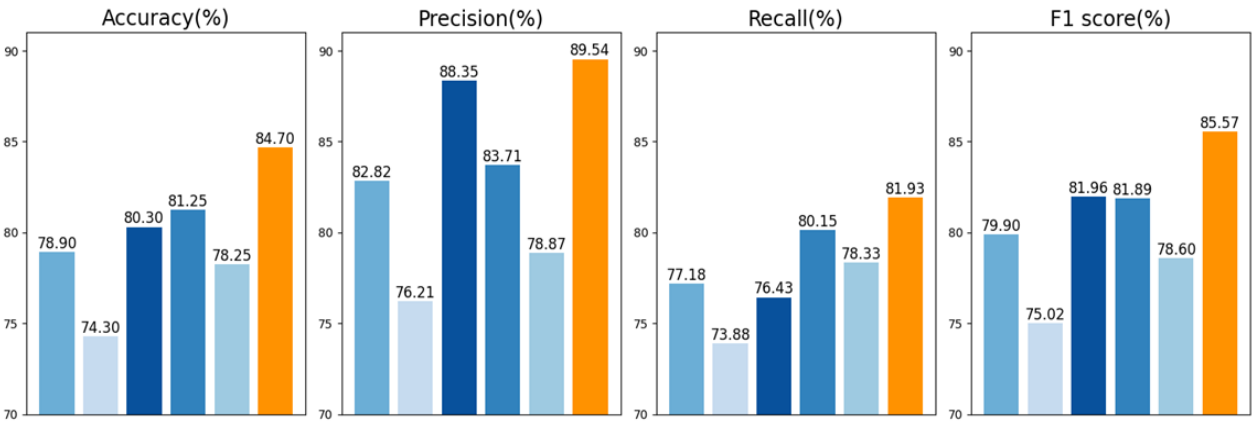

그림 8. 기본 모델과 제안한 모델 간 성능 비교
Fig. 8. Comparison between baseline and proposed model

Ⅴ. Conclusions

본 연구에서는 이미지 조작 검출을 위해 특징 강화를 적용한 앙상블 기반의 이미지 조작 검출 시스템을 제안하였다. 이미지 조작은 그 형태가 수없이 많고, 끊임없이 새롭게 생산되어 타인에게 피해를 줄 수 있으므로 다양한 형태에 대한 성능과 새로운 형태에 대한 빠른 적용이 필요하다. 본 연구는 효과적인 이미지 조작을 검출할 때의 다양성과 신규성을 효과적으로 반영하기 위해 실제 다양한 조작 형태가 반영된 최신의 DISC21 데이터를 활용하였다. 이미지의 조작 특징을 강화하기 위한 기법으로 적응적 임계화와 캐니 에지 검출이 적용되었으며, 다중 모델의 장점을 효과적으로 활용한 소프트 보팅 앙상블을 통해 다양성과 강건함을 가진 검출 시스템을 제안하였다. 각 모델은 이미지 조작 검출 분야에서도 높은 성능을 보이는 CNN 및 Transformer 기반의 9가지 모델을 활용하였고, 앙상블 프레임워크에는 가장 우수한 성능을 낸 VGG16, VGG19, EfficientNet-b0, EfficientNet-b7, Swin Transformer 모델을 채택하였다. 실험을 통해 정확도 84.70%, 정밀도 89.54%, 재현율 81.93%, F1-score 85.57%를 달성하였으며, 단일 기본 모델의 성능과 비교했을 때 최대 정확도 10.4%, 정밀도 13.33%, 재현율 8.05%, F1 점수 10.55%가 향상되었다. 제안된 시스템은 방대한 새로운 조작 유형을 빠르게 필터링하여 의도된 이미지 조작과 잠재된 피해를 사전에 방지하고, 끊임없이 생성되는 새로운 유형의 조작에 대해 쉽게 확장이 가능한 프레임워크를 제안했다는 점에서 본 연구의 의의를 찾을 수 있다.

그러나 육안으로도 조작을 판별하기 어려운 이미지가 존재하여 조작 경계만을 강화하는 것만으로 검출 성능 확보가 어려운 경우가 존재한다는 것이 연구의 한계라고 볼 수 있다. 향후 연구 과제로는 조작 경계가 자연스러운 경우에서의 검출 성능을 높이기 위해 추출된 특징을 결합하여 학습하는 등 새로운 정보를 학습하고, 이미지 조작 유형을 각각 분리하여 충분한 추가 데이터를 학습하여 각 유형에 최적화된 모델로 앙상블을 구성하여 더욱 강건하고, 확장성을 높이는 연구를 수행해 보고자 한다.

References

[1] S. S. Ali, I. I. Ganapathi, N.-S. Vu, S. D. Ali, N. Saxena, and N. Werghi, "Image Forgery Detection Using Deeplearning by Recompressing Images", Electron., Vol. 11, No. 3, pp. 403, Feb. 2022. https://doi.org/10.3390/electronics11030403.

[2] S. Tyagi and D. Yadav, "A detailed analysis of image and video forgery detection techniques", The Visual Computer, Vol. 39, No. 3, pp. 813-833, Jan. 2022. https://doi.org/10.1007/s00371-021-02347-4.

[3] F. Z. Mehrjardi, A. M. Latif, M. S. Zarchi, and R. Sheikhpour, "A survey on deep learning-based image forgery detection", Pattern Recognit., Vol. 144, Dec. 2023. https://doi.org/10.1016/j.patcog.2023.109778.

[4] B. Shah, D. Shah, S. Thakar, S. Shah, and S. Dhage, "Image Manipulation Detection Using Error Level Analysis", 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, pp. 1-6, Aug. 2023. https://doi.org/10.1109/ASIANCON58793.2023.10270614.

[5] K. R. Revi, M. Wilscy, and R. Antony, "Portrait photography splicing detection using ensemble of convolutional neural networks", Journal of Intelligent & Fuzzy Systems, Vol. 41, No. 5, pp. 5347-5357, Jan. 2021. https://doi.org/10.3233/JIFS-189857.

[6] K. B. Meena and V. Tyagi, "Image Forgery Detection: Survey and Future Directions", Data, Engineering and Applications, pp. 163-194, Apr. 2019. https://doi.org/10.1007/978-981-13-6351-1_14.

[7] E. U. H. Qazi, T. Zia, and A. Almorjan, "Deep Learning-Based Digital Image Forgery Detection System", Appl. Sci., Vol. 12, No. 6, Mar. 2022. https://doi.org/10.3390/app12062851.

[8] N. T. Pham and C. S. Park, "Toward Deep-Learning-Based Methods in Image Forgery Detection: A Survey", IEEE Access, Vol. 11, pp.

11224-11237, Feb. 2023. https://doi.org/10.1109/ACCESS.2023.3241837.

[9] M. Zanardelli, F. Guerrini, R. Leonardi, and N. Adami, "Image forgery detection: a survey of recent deep-learning approaches", Multimedia Tools and Applications, Vol. 82, No. 12, pp. 17521-17566, Oct. 2023. https://doi.org/10.1007/s11042-022-13797-w.

[10] H. Li and J. Huang, "Localization of Deep Inpainting Using High-Pass Fully Convolutional Network", 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 8300-8309, Oct. 2019. https://doi.org/10.1109/ICCV.2019.00839.

[11] N. K. Hebbar and A. S. Kunte, "Transfer Learning Approach for Splicing and Copy-Move Image Tampering Detection", ICTACT Journal on Image and Video Processing, Vol. 11 , No. 4, pp. 2447-2452, May 2021.

[12] N. Goel, S. Kaur, and R. Bala, "Dual branch convolutional neural network for copy move forgery detection", IET Image Processing, Vol. 15, No. 3, pp. 656-665, Feb. 2021. https://doi.org/10.1049/ipr2.12051.

[13] B. T. Hammad, I. T. Ahmed, and N. Jamil, "An Secure and Effective Copy Move Detection Based on Pretrained Model", 2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, pp. 66-70, Jul. 2022. https://doi.org/10.1109/ICSGRC55096.2022.9845141.

[14] K. D. Kadam, S. Ahirrao, K. Kotecha, and S. Sahu, "Detection and Localization of Multiple Image Splicing Using MobileNet V1", IEEE Access, vol. 9, pp. 162499-162519, Nov. 2021. https://doi.org/10.1109/ACCESS.2021.3130342.

[15] M. N. Abbas, M. S. Ansari, M. N. Asghar, N. Kanwal, T. O'Neill, and B. Lee, "Lightweight Deep Learning Model for Detection of Copy-Move Image Forgery with Post-Processed Attacks", 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, pp. 125-130, Jan. 2021. https://doi.org/10.1109/SAMI50585.2021.9378690.

[16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv Prepr, Apr. 2015. https://doi.org/10.48550/arXiv.1409.1556.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778, Jun. 2016. https://doi.org/10.1109/CVPR.2016.90.

[18] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", Proceedings of the 36th International Conference on Machine Learning, PMLR, Long Beach, California, USA, Vol. 97, pp. 6105-6114, May 2019.

[19] A. Dosovitskiy, et al., "An image is worth 16x16 words: Transformers for image recognition at scale", arXiv Prepr, Jun. 2021. https://doi.org/10.48550/arXiv.2010.11929.

[20] N. H. Rajini, "Image forgery identification using convolution neural network", International Journal Recent Technology Engineering, Vol. 8, No. 1, pp. 311-320, Jun. 2019.

[21] Abhishek and N. Jindal, "Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation", Multimed. Tools Applications, Vol. 80, No. 3, pp. 3571-3599, Jan. 2021. https://doi.org/10.1007/s11042-020-09816-3.

[22] S. Kumar, S. K. Gupta, M. Kaur, and U. Gupta, "VI-NET: A hybrid deep convolutional neural network using VGG and inception V3 model for copy-move forgery classification", Journal of Visual Communication and Image Representation, Vol. 89, Nov. 2022. https://doi.org/10.1016/j.jvcir.2022.103644.

[23] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD - New database for copy-move forgery detection", Proceedings ELMAR-2013, Zadar, Croatia, pp. 49-54, Sep. 2013.

[24] M. Aria, M. Hashemzadeh, and N. Farajzadeh, "QDL-CMFD: A Quality-independent and deep Learning-based Copy-Move image forgery detection method", Neurocomputing, Vol. 511, pp. 213-236, Oct. 2022. https://doi.org/10.1016/j.neucom.2022.09.017.

[25] S. Ganguly, A. Ganguly, S. Mohiuddin, S. Malakar, and R. Sarkar, "ViXNet: Vision Transformer with Xception Network for deepfakes based video and image forgery detection", Expert Systems with Applications, Vol. 210, pp. 118423, Jul. 2022. https://doi.org/10.1016/j.eswa.2022.118423.

[26] H. Kashiani, S. M. Sami, S. Soleymani, and N. M. Nasrabadi, "Robust Ensemble Morph Detection with Domain Generalization", 2022 IEEE International Joint Conference on Biometrics (IJCB), Abu Dhabi, United Arab Emirates, pp. 1-10, Oct. 2022. https://doi.org/10.1109/IJCB54206.2022.10007929.

[27] A. Hashmi, S. A. Shahzad, W. Ahmad, C. W. Lin, Y. Tsao, and H.-M. Wang, "Multimodal Forgery Detection Using Ensemble Learning", 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, pp. 1524-1532, Nov. 2022. https://doi.org/10.23919/APSIPAASC55919.2022.9980255.

[28] E. Tasci, C. Uluturk, and A. Ugur, "A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection", Neural Computing Applications, Vol. 33, No. 22, pp. 15541-15555, Jun. 2021. https://doi.org/10.1007/s00521-021-06177-2.

[29] M. Douze, et al., "The 2021 image similarity dataset and challenge", arXiv Prepr, Feb. 2022. https://doi.org/10.48550/arXiv.2106.09672.

[30] B. Thomee, et al., "YFCC100M: The new data in multimedia research", Communications of the ACM, Vol. 59, No. 2, pp. 64-73, Jan. 2016. https://doi.org/10.1145/2812802.

[31] B. Dolhansky, et al., "The DeepFake Detection Challenge (DFDC) Dataset", arXiv Prepr, Oct. 2020. https://doi.org/10.48550/arXiv.2006.07397.

[32] C. Hazirbas, J. Bitton, B. Dolhansky, J. Pan, A. Gordo, and C. C. Ferrer, "Towards Measuring Fairness in AI: The Casual Conversations Dataset", IEEE Transactions on Biometrics, Behavior, and Identity Science, Vol. 4, No. 3, pp. 324-332, Jul. 2022. https://doi.org/10.1109/TBIOM.2021.3132237.

[33] D. Bradley and G. Roth, "Adaptive Thresholding using the Integral Image", Journal of Graphics Tools, Vol. 12, No. 2, pp. 13-21, Jan. 2007. http://dx.doi.org/10.1080/2151237X.2007.10129236.

[34] A. S. Methil, "Brain Tumor Detection using Deep Learning and Image Processing", 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, pp. 100-108, Mar. 2021. https://doi.org/10.1109/ICAIS50930.2021.9395823.

[35] V. Ribeiro, et al., "Brazilian Mercosur License Plate Detection: A Deep Learning Approach Relying on Synthetic Imagery", 2019 IX Brazilian Symposium on Computing Systems Engineering (SBESC), Natal, Brazil, pp. 1-8, Nov. 2019. https://doi.org/10.1109/SBESC49506.2019.9046091.

[36] J. Canny, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 6, pp. 679-698, Nov. 1986. https://doi.org/10.1109/TPAMI.1986.4767851.

[37] 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 (CVPR), Salt Lake City, UT, USA,

pp. 4510-4520, Jun. 2018. https://doi.org/10.1109/CVPR.2018.00474.

[38] Z. Liu, et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows", 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 9992-10002, Oct. 2021. https://doi.org/10.1109/ICCV48922.2021.00986.

[39] J. Park, S. Lee, Y.-J. Ban, G. Min, and J. Kim, "Diabetic Foot Ulcer Diagnosis using Deep Learning", The Journal of Korean Institute of Information Technology, Vol. 22, No. 2, pp. 35-42, Feb. 2024. https://doi.org/10.14801/jkiit.2024.22.2.35.

[40] W.-Y. Baek and S.-G. Kang, "Ship Classification Method using Two-Stage CNN Model", The Journal of Korean Institute of Information Technology, Vol. 21, No. 8, pp. 203-210, Aug. 2023. https://doi.org/10.14801/jkiit.2023.21.8.203.

[41] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review", Engineering Applications of Artificial Intelligence, Vol. 115, Oct. 2022. https://doi.org/10.1016/j.engappai.2022.105151.

[42] N. Qian, "On the momentum term in gradient descent learning algorithms", Neural networks, Vol. 12, No. 1, pp. 145-151, Jan. 1999. https://doi.org/10.1016/S0893-6080(98)00116-6.

[43] L. Bondi, S. Lameri, D. Guera, P. Bestagini, E. Delp, and S. Tubaro, "Tampering Detection and Localization Through Clustering of Camera-Based CNN Features", 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, pp. 1855-1864, Jul. 2017. https://doi.org/10.1109/CVPRW.2017.232.

[44] D. Mallick, M. Shaikh, A. Gulhane, and T. Maktum, "Copy Move and Splicing Image Forgery Detection using CNN", ITM Web of Conferences, Vol. 44, pp. 03052, May 2022. https://doi.org/10.1051/itmconf/20224403052.

저자소개

길 현 선 (Hyunsun Kil)

2021년 2월 : 가톨릭대학교 컴퓨터정보공학부(공학사)
2024년 2월 : 고려대학교 산업경영공학과(공학석사)
관심분야 : 인공지능, 컴퓨터 비전, 조작 탐지

김 지 호 (Jiho Kim)

2015년 8월 : 서울과학기술대학교 글로벌융합산업공학과(공학사)
2015년 9월 ~ 현재 : 고려대학교 산업경영공학과 석박사통합과정
관심분야 : 인공지능, 자연어처리, 비즈니스 인텔리전스

이 홍 철 (Hongchul Lee)

1983년 : 고려대학교 산업공학부(공학사)
1988년 : University of Texas, Arlington Industrial Engineering (M.S.)
1993년 : Texas A&M University, Industrial Engineering(Ph.D.)
1996년 ~ 현재 : 고려대학교 산업경영공학과 교수
관심분야 : 인공지능, 생산·물류 시스템, 시뮬레이션