ISSN : 1976－5304 / (Online)2713－931X
디지털포렌식연구
제19권 제4호(2025.08) 116－131쪽
(사단법인)한국디지털포렌식학회

DOI : 10.22798 / kdfs.2025.19.4.116
Journal of Digital Forensics
Vol.19 No.4(Aug 2025) pp.116－131
Korean Digital Forensics Society

눈·코·입 영역 강조 가중치를 활용한 ViT 기반 딥페이크 탐지 기법 연구

엄 기 원*, 김 동 수**, 오 태 우***, 장 한 얼****
국립한밭대학교 컴퓨터공학과 (학부생)*, 국립한밭대학교 컴퓨터공학과 (대학원생)**,
국가보안기술연구소 (책임 연구원)***, 국립한밭대학교 컴퓨터공학과 (교수)****

ViT－based Deepfake Detection via Regional Emphasis Weights on Eyes, Nose, and Mouth

Ki－won Eom*, Dong－Su Kim**, Tae－Woo Oh*** Haneol Jang****
Dept. of Computer Engineering, Hanbat National University (Undergraduate)*, (Graduate Student)**,
National Security Research Institute (Principal Researcher)***,
Dept. of Computer Engineering, Hanbat National University (Professor)****

■ 요약 ■

딥페이크 기술이 빠르게 고도화됨에 따라, 이를 악용한 가짜뉴스 유포나 디지털 성범죄와 같은 사회적 문제가 심각해지고 있다. 따라서 이러한 딥페이크 콘텐츠를 정확하게 탐지하고 대응하기 위한 기술이 중요한 과제로 떠오르고 있다. 기존 연구들은 CNN 혹은 ViT 기반의 모델을 사용해 딥페이크 탐지를 진행하였다. 하지만 CNN은 전역적인 정보, ViT는 국소적인 정보를 상대적으로 탐지하지 못하는 문제점이 존재한다. 본 연구에서는 ViT 기반 모델의 한계점인 국소 정보 탐지를 보완하기 위해 MediaPipe를 이용해 얼굴 랜드마크를 추출하고, ViT의 패치 임베딩 단계에 눈·코·입 영역별 학습 가능한 가중치를 부여하는 Region Weight 모듈 도입을 제안한다. 이를 통해 ViT 모델이 전역 정보 뿐 아니라 딥페이크 기법에서 주로 발생하는 눈·코·입 주변의 미세한 변화를 효과적으로 학습하도록 한다. FaceForensics＋＋ 데이터셋을 활용해 실험한 결과, Region Weight Module 사용 전 ViT 대비 ACER는 7.94%, AUC는 약 3% 개선되었다. 본 논문의 제안 기법은 발전하는 딥페이크 생성 기법에서 발생하는 주요 특징을 효과적으로 포착하며, 향후 더욱 정교하고 은밀해지는 딥페이크 기법에 대해서도 안정적으로 대응할 수 있는 기반 기술로 활용될 수 있을 것으로 기대된다.

● 주제어 : 딥페이크 탐지, 비전트랜스포머 모델, 랜드마크, 패치 가중치, 디지털포렌식

■ ABSTRACT ■

As deepfake technology rapidly advances, its misuse for the dissemination of fake news and digital sex crimes has become a serious social issue. Consequently, developing technologies capable of accurately detecting and countering such malicious content has emerged as a critical challenge. Prior work has employed CNN－ or ViT－based models for deepfake

※ 이 논문은 한국방송미디어공학회 2025년 하계학술대회에 발표된 학술발표논문을 확장하였음.
■ 투고일 : 2025.07.22. 심사게시일 : 2025.07.22. 게재확정일 : 2025.08.28.
■ 제1저자(First Author) : Ki－won Eom, Dong－Su Kim (Email : 20231203@edu.hanbat.ac.kr, 30251264@edu.hanbat.ac.kr)
■ 교신저자(Corresponding Author) : Tae－Woo Oh, Haneol Jang (Email : twoh@nsr.re.kr, hejang@hanbat.ac.kr)

detection. A limitation exists, however, in that CNNs are relatively weak at capturing global context, and ViTs are comparatively poor at detecting local information. To address the ViT's weakness in local information sensing, we first extract facial landmarks using MediaPipe, then introduce a Region Weight module that assigns learnable weights to the eye, nose, and mouth regions during the ViT's patch embedding stage. This enhancement allows the model to learn not only global patterns but also the subtle artifacts around the eyes, nose, and mouth that are characteristic of deepfake manipulations. Experimental results on the FaceForensics++ dataset demonstrate that, relative to the baseline ViT, incorporating the Region Weight Module yields a 7.94% reduction in ACER and an approximately 3% increase in AUC. The proposed method effectively captures the salient features emerging from evolving deepfake generation techniques and is expected to provide a robust foundation for countering ever more sophisticated and covert deepfake methods in the future.

● Key Words : Deepfake Detection, Vision Transformer (ViT) Model, Facial Landmark, Patch Weight, Digital Forensics

Ⅰ. 서 론

딥페이크(Deepfake)는 deep learning과 fake의 합성어로, 실제와 구분할 수 없을 정도로 정교하게 만든 가짜 미디어를 일컫는다. 오늘날 생성적 적대 신경망(Generative Adversarial Network: GAN)[1]과 확산 모델(Diffusion Models)[2]의 비약적인 발전으로 인해, 딥페이크 기술은 나날이 정교해지고 있다. 이러한 딥페이크 기술은 엔터테인먼트, 미디어 등 다양한 분야에서 긍정적인 활용 가능성을 보여주는 한편, 허위 정보 유포, 디지털 성범죄, 금융 사기와 같은 심각한 사회적·법적 문제를 야기하고 있다. 예컨대 한 장의 얼굴 사진만으로도 그 얼굴에 어울리는 가짜 음성을 생성해 음성 인증을 우회할 수 있으며[3], 이처럼 얼굴·음성·영상 데이터를 활용한 다양한 공격 기법의 발전으로 인해 단순한 허위 영상 제작을 넘어, 심각한 범죄에 악용되는 사례가 증가하고 있어 딥페이크 탐지법의 발 빠른 발전 필요성이 대두되고 있다.

딥페이크 탐지 연구 초기에는 주로 CNN(Convolutional Neural Network)을 기반으로 텍스처 불일치나 경계선 흐림과 같은 생성 과정에서 발생하는 세부적인 왜곡 흔적(artifact)을 포착하는 방법이 제안되었다. 대표적인 예로는 XceptionNet[4], ResNet[5] 계열을 활용한 CNN 모델들이 있으며, 합성 흔적을 효과적으로 감지하여 높은 분류 성능을 보여주었다. 하지만 고정된 수용 영역(Receptive Field)과 계층적 구조로 인해 얼굴의 구조적 패턴을 전역적인 관점에서 반영하는 데 한계를 드러냈다. 이와 달리 Vision Transformer(ViT)[6]를 통한 딥페이크 탐지 방식은 입력 이미지를 다수의 패치로 분할하고, Self-attention을 통해 전역적 패턴을 효과적으로 학습할 수 있다. 하지만 ViT 기반 탐지 모델은 얼굴 전체의 불일치 패턴을 효과적으로 포착하는 반면, 패치 분할 크기와 위치 정보 의존성으로 인해 상대적으로 미세한 영역의 왜곡 및 합성 흔적을 놓치기 쉽다는 한계가 존재한다.

본 논문에서는 딥페이크 생성 및 합성 과정에서 눈·코·입 영역에 합성 흔적이 많이 발생한다는 점을 이용하여, 기존의 ViT 모델이 가진 한계를 극복하기 위해 눈·코·입 영역이 해당하는 패치에 더 가중치를 주도록 유도하는 Region Weight Module을 제안한다. 우선, 얼굴 검출 딥러닝 모델인 MTCNN(Multi-task Cascaded Convolutional Networks)을 이용한 얼굴 부위 Crop 및 Resize로 전처리한 데이터들에서 얼굴의 눈·코·입 영역에 해당하는 랜드마크 좌표들을 추출한 후, 좌표가 속하는 패치와 매핑한다. 이후 생성된 가

중치 맵 정보를 바탕으로, ViT의 encoder를 통한 연산으로 해당 부위에 대한 가중치를 부여할 수 있게 된다. 본 논문에서 제안한 기법을 통해 기존 ViT의 한계점인 미세 왜곡 포착을 보완하며 핵심 얼굴 부위의 특징을 더욱 뚜렷하게 강조하는 CNN의 강점을 흡수할 수 있게 된다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 딥페이크 탐지 기법 연구 동향 및 딥페이크 생성물의 왜곡 흔적 정보, 그리고 일반화 실험의 기존 사례에 대하여 종합적으로 고찰한다. 3장에서는 본 논문의 제안 기법인 Region Weight Module의 원리 및 구현 세부사항, 모델 입력 구조 과정을 상세히 서술한다. 4장에서는 사용한 데이터셋 및 실험 설정, 실험 결과 분석을 통해 본 논문의 제안 기법의 우수성을 입증하며, 5장에서는 본 연구의 결론 및 향후 연구 방향을 논의한다.

II. 관련 연구

2.1. 기존 딥페이크 탐지법

딥페이크 탐지 연구는 CNN 기반 방법에서 시작되었다. 대표적으로 Xception·ResNet 계열 모델은 깊은 합성곱(convolution) 필터를 통해 이미지의 텍스처 불일치·경계선 흐림 등을 효과적으로 포착해 높은 정확도를 달성했다. 그러나 CNN은 고정된 수용 영역 탓에 얼굴 구조를 전역적 관점에서 해석하는 데 한계가 있었다. 이 한계를 보완하기 위해 특정 특징을 강화하는 Face X−Ray[7], F3−Net[8] 등이 도입되었다. Face X−Ray는 원본에서 합성 얼굴로 블렌딩될 때 생기는 경계면을 단일 회색 조 이미지로 추출하여 다양한 조작 방식에 범용적으로 대응하는 방법이다. 이러한 방식을 통해 Unseen Domain 데이터에도 좋은 일반화 성능을 입증했다. 또한 F3−Net은 주파수 영역 시각화(Frequency−aware Decomposition, FAD)와 국소 주파수 통계(Local Frequency Statistics, LFS)를 두 스트림으로 학습해 압축·노이즈에 의한 아티팩트까지 검출 가능한 주파수 기반 탐지기를 구현했다. 또한, CBAM[9]과 같이 중간 특징에서 중요도를 자동 학습하는 CNN의 채널 어텐션이 제안되었다. 하지만 CBAM은 결과가 레이어별 단일 채널의 2D 어텐션 맵으로 모든 채널에 일괄 적용되기 때문에, 특정 패치만 선택적으로 증폭·억제하는 정밀 제어에 한계가 있다. 최근에는 ViT 구조가 주목받고 있다. ViT는 입력 이미지를 패치로 분할 후, 어텐션으로 전역 문맥을 학습함으로써 CNN 대비 일반화 성능이 크게 향상되었다. 다만 모든 패치를 전역적으로 다루는 특성상, 눈·코·입같이 딥페이크 합성 흔적이 집중되는 국소 영역의 왜곡을 상대적으로 약하게 학습하는 한계가 존재한다. 실제로 딥페이크는 얼굴 전체가 아닌 특정 부위에만 국소적으로 조작이 이루어지는 경우가 많으며, 이로 인해 모델이 중요 영역과 그렇지 않은 영역을 구분 없이 처리할 경우, 탐지 정확도가 저하될 수 있다. 이를 보완하려는 시도로, 패치의 중요도를 학습으로 유도하는 ViT 계열 개량 모델인 FakeFormer[10]가 존재한다. 그러나 FakeFormer는 의미 기반 위치 정보를 직접 주입하지는 않는다는 제약이 존재한다. 이에 본 논문은 눈·코·입 위치의 패치를 선별한 뒤, 해당 패치 임베딩에 학습 가능한 가중치를 적용하는 패치 어텐션 메커니즘이 적용된 모듈인 Region Weight Module을 도입한다. 이렇게 강조된 패치 임베딩은 ViT의 패치 기반 처리와 자연스럽게 결합되어, 국소적 합성 흔적에 대한 민감도를 효과적으로 높인다.

2.2. 딥페이크 합성물의 Artifact

딥페이크 영상은 합성 기법이나 데이터셋이 다르더라도 여러 연구에서 몇 가지 공통적인 시각적 결함을 드러내는 것으로 보고된다[11]. 대표적으로 경계 불일치로 인해 눈가·코·입술 주위 경계선이 흐릿해지거나 다중 경계가 형성되고, 텍스처 저해상도 현상으로 피부 모공·주름 같은 고주파 정보가 손실되어 왁스 인형처럼 보이는 느낌이 나타나기도 한다. 이러한 흔적들은 프레임 간 일관성이 떨어질 때 미세한 깜빡임으로 이어지기도 하여, 주파수·시공간 도메인 기반 딥페이크 탐지 모델이 활용할 수 있는 중요한 단서가 될 수 있다. <Figure 1>은 딥페이크 이미지와 원본 이미지 간의 눈·코·입술 부위를 비교한 예로, 경계선 흐림과 픽셀 왜곡 발생이 뚜렷하게 나타남을 보여준다. 이러한 왜곡 흔적은 합성 주체·조도·압축률이 달라지는 경우에도 필연적으로 발생하므로, 탐지 모델이 집중해야 할 핵심 단서임을 시사한다. 본 논문에서는 이러한 연구들을 토대로 눈·코·입의 세 부위에 보다 집중하도록 눈·코·입의 좌표가 포함되는 패치에 가중치를 주는 방식으로 Region Weight Module을 고안하였다.

〈Figure 1〉 Eyes, Nose, and Mouth boundary artifacts between real and fake images

2.3 딥페이크 일반화(Generalization) 실험

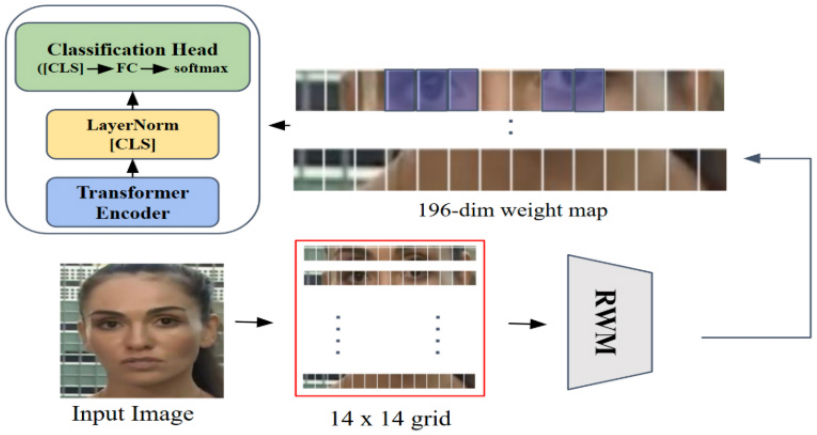
딥페이크 탐지 모델이 실제 환경에서 좋은 성능을 가졌다는 신뢰성을 입증하려면, 학습에 사용되지 않은 합성 기법에서도 일관된 성능을 내야 한다. 최근 얼굴 안티스푸핑 분야에서는 데이터셋 간 도메인 차이를 극복하기 위해 Leave−one−out 평가를 표준으로 삼고 있다. 예를 들어 Kim et al.[12]는 네 개 데이터셋(OULU, CASIA, Idiap, MSU−MFSD) 중 세 개로 학습하고 남은 하나로만 테스트하는 Zero−Shot Domain Generalization(ZSDG) 프로토콜을 제시하여, 여러 도메인으로 학습한 뒤 학습에 사용하지 않은 도메인에서 검증하는 실험 설계가 딥페이크 탐지 모델의 일반화 연구에 필요함을 입증했다.

본 논문은 이러한 프로토콜을 참고하여 합성 기법을 각각의 도메인으로 설정하여 실험을 확장하였다. 이러한 추가 실험 설계를 통해 일반화 실험을 효과적으로 수행함으로써, Region Weight Module이 적용된 ViT가 학습에 포함되지 않은 합성 기법에서도 기존 ViT 모델보다 높은 탐지 성능을 유지함을 입증하고자 하였다.

Ⅲ. 제안 기법

본 논문에서는 ViT 모델에 Region Weight 모듈을 결합해 얼굴 주요 부위인 눈·코·입 영역이 속한 패치에 학습 가능한 가중치를 적용한다. <Figure 2>는 Region Weight 모듈을 적용한 ViT 기반 전체 파이프라인이다.

〈Figure 2〉 Region Weight Module Applied ViT-based Deepfake Detection Pipeline



3.1 랜드마크 추출

입력 이미지는 224×224픽셀로 Resize한 뒤, MediaPipe를 이용해 눈·코·입에 해당하는 랜드마크 세트를 추출한다. MediaPipe Face Mesh[13]는 단일 RGB 영상에서 468개 3D 얼굴 랜드마크를 실시간으로 예측하는 경량 딥러닝 모델이다. 본 연구는 <Figure 3>과 같이 MediaPipe를 이용해 눈·코·입과 관련된 세밀한 좌표를 정확히 추출하였으며, 양쪽 눈 동공, 코끝, 입술 양끝의 5개 대표 지점을 담은 5-point 세트와 부위의 모든 세부 지점(133개)을 포함한 Full-point 세트를 각각 구성하여 실험을 진행한 후, 두 세트의 성능 차이를 비교하였다.

〈Figure 3〉 Example of landmark extraction using MediaPipe
(left: 5-point set; right: full-point set)

5-point Set **Full-point Set**

3.2 Region Weight Module(RWM)

본 연구에서는 이미지를 16×16 크기의 패치로 분할하여 한 행에 14개, 총 196개의 패치를 생성한 후, MediaPipe로 추출한 얼굴 랜드마크의 좌표를 해당 패치 인덱스에 매핑함으로써 부위별 위치 정보를 반영하였다. 이후 눈·코·입 세 부위에 대한 이진마스크($M_{r,i}$)를 생성하고, 이를 기반으로 강조 가중치를 적용하는 Region Weight Module(RWM)을 구성하였다. 각 패치 i에 대한 최종 가중치 W_i는 수식(1)과 같이 정의된다.

$$W_i = w_{default} + \sum_{r \in eye, nose, mouth} (w_r - w_{default}) M_{r,i},\ i = 1, \cdots N \tag{1}$$

여기서 $M_{r,i}$는 패치 i가 부위 r에 속하면 1, 아니면 0인 이진 지시 함수이다. 예로 들어 눈 부위의 경우, 눈 부위의 학습 파라미터 w_{eye}와 나머지 패치에 적용되는 기본 가중치 $w_{default}$의 차이를 구한 후, 이진 마스크 함수를 곱해준다. 코·입에도 동일한 절차를 거치고, 최종적으로 수식(1)을 통해 각 패치의 위치에 따라 해당 부위의 가중치가 적용되며, 해당 영역에 속하지 않는 패치는 기본 가중치를 그대로 유지하게 된다. 본 연구에서는 RWM의 가중치 생성 방식을 Fixed, Single−learnable, Multi−learnable의 세 가지 모드로 정의하였다. 먼저, Fixed 모드는 모든 강조 가중치를 고정된 상수로 설정하여 학습 과정 중 변경되지 않도록 한다. 이 방식은 모델의 학습 안정성 측면에서는 유리하지만, 최적의 강조 비율을 자동으로 찾을 수 없다는 한계가 있다. 둘째, Single−learnable 모드는 눈·코·입에 대해 하나의 공통된 학습 가능한 파라미터를 도입하여, 학습을 통해 전체 부위에 적용할 최적의 강조 값을 자동으로 조정하도록 한다. 마지막으로, Multi−learnable 모드는 각 부위별로 개별적인 학습 가능한 가중치 파라미터를 설정하여, 각 얼굴 부위의 중요도를 독립적으로 미세하게 조절할 수 있도록 한다. 이처럼 Single 및 Multi−learnable 모드에서 사용되는 부위별 강조 가중치는 학습 가능한 파라미터로 정의되며, 손실 함수의 Gradient를 통해 전체 분류 성능을 향상시키는 방향으로 자동 조정된다. 이를 통해 모델은 특정 부위에 딥페이크 흔적이 집중되는 경향이 있음을 데이터 기반으로 학습하며 부위별로 서로 다른 딥페이크 흔적 특성을 포착할 수 있다는 장점이 있으며, 본 연구에서는 해당 모드를 주요 실험 대상으로 삼고 다른 방식과의 성능을 비교하였다.

3.3 최종 모델 구조

입력 이미지는 총 196개 패치로 분할되며, 각 패치에 대해 임베딩 벡터를 추출한다. 이후 RWM에서 계산된 패치 어텐션 가중치를 각 임베딩에 요소별 곱셈(element−wise multiplication) 방식으로 적용하여, 눈·코·입 위치의 패치가 선택적으로 강조된 패치 임베딩이 생성된다. 여기서 RWM은 눈·코·입 위치가 속한 패치 임베딩에 학습 가능한 가중치를 부여함으로써, 패치 단위의 중요도를 명시적으로 주입하는 패치 어텐션으로 작동한다. 이렇게 강조된 패치 임베딩은 ViT 모델의 Transformer Encoder에 입력되며, Self−attention 메커니즘을 통해 패치 간의 관계를 학습한다. 이때 각 패치 임베딩과 함께 특수 토큰인

CLS(Classification Token)가 추가되며, 이는 전체 입력 시퀀스를 대표하는 벡터로 사용된다. Transformer Encoder를 통과한 후에는 CLS 토큰에 해당하는 출력 벡터를 기반으로 FC(Fully Connected) 레이어와 Softmax 분류기를 거쳐 해당 이미지의 Real/Fake 여부에 대한 이진 분류를 수행한다.

Ⅳ. 실 험

4.1. 데이터셋

본 연구의 실험은 딥페이크 탐지 분야에서 널리 활용되는 FaceForensics++ 데이터셋[14]을 기반으로 진행하였다. FaceForensics++는 뮌헨 공과대학교(TU Munich)에서 구축한 고해상도 얼굴 합성 영상 데이터셋으로, 실제 방송 인터뷰 영상에서 수집된 얼굴 클립을 기반으로 다양한 합성 기법을 적용한 변조 영상으로 구성된다. 이 데이터셋은 다양한 연구 및 관련 논문에서 표준 벤치마크로 사용되고 있다. 본 연구에서는 FaceForensics++로부터 원본 비디오 1,000개와 DeepFakes, Face2Face, FaceSwap, FaceShifter, NeuralTextures의 다섯 가지 합성 기법으로 생성된 비디오 각 1,000개씩을 포함하여 총 6,000개의 영상을 활용하였다. 각 기법의 주요 특징은 다음과 같다. DeepFakes는 인코더−디코더 구조를 활용하여 두 얼굴 간의 잠재 표현을 학습한 후, 소스 얼굴을 타겟 얼굴로 교체하는 오픈소스 기반 얼굴 변환 기법이다. Face2Face는 소스 영상에서 추출한 얼굴 표정을 타겟 영상의 얼굴에 실시간으로 반영하여 표정을 조작하는 방식이다. FaceSwap은 얼굴 영역을 정렬한 후 직접적으로 교체하는 얼굴 스왑 기법이다. FaceShifter는 얼굴의 정체성을 유지하면서도 고해상도 품질을 달성하는 딥러닝 기반의 고정밀 얼굴 합성 기법이다. 마지막으로, NeuralTextures는 3D 얼굴 모델 기반 텍스처와 신경망을 통합하여 사실적인 얼굴 영상을 합성하는 신경 렌더링 기반 기법이다. 각 영상으로부터는 전체 프레임 중 5등분된 시점의 프레임만을 추출하여 사용하였다. 예를 들어, 한 영상이 100프레임으로 구성된 경우 1, 20, 40, 60, 80번째 프레임이 선택된다. 이러한 방식으로 총 6,000개의 영상으로부터 30,000장의 프레임 이미지를 확보하였으며, 이후 MTCNN 라이브러리를 사용하여 각 프레임 내 얼굴 영역을 검출하고 crop하여 최종 실험용 얼굴 이미지로 활용하였다. 이미지들은 다양한 해상도를 가지고 있기 때문에 224×224 Resize를 일괄 적용하였다. 최종적으로 사용된 이미지의 예시는 <Figure 4>에 제시되어 있으며, 왼쪽 상단부터 차례대로 원본, DeepFakes, Face2Face, FaceShifter, FaceSwap, NeuralTextures 순으로 나열되어 있다.

〈Figure 4〉 Example facial images extracted from the FaceForensics++ dataset

4.2. 평가 지표

본 연구에서는 딥페이크 탐지 모델의 성능을 정량적으로 평가하기 위해, 주요 평가 지표로 ACER (Average Classification Error Rate)를, 보조 지표로 AUC (Area Under the Curve)를 사용하였다. 사용된 FaceForensics++ 데이터셋은 Real 데이터보다 Fake 데이터가 상대적으로 많아 클래스 간 불균형이 존재한다. 따라서 정확도(Accuracy)는 분류 성능을 부정확하게 평가할 가능성이 있어 본 연구에서는 정확도를 주요 평가 지표로 사용하지 않았다. 대신, 불균형 상황에서 실제와 조작 샘플 모두에 대한 오탐과 누락을 균형 있게 평가할 수 있는 ACER를 핵심 지표로 채택하였다.

4.2.1 평균 오분류율(ACER)

ACER는 두 가지 하위 지표인 APCER(Attack Presentation Classification Error Rate)와 BPCER(Bona Fide Presentation Classification Error Rate)의 평균으로 정의되며, 딥페이크 및 얼굴 스푸핑 탐지 분야에서 핵심 지표로 널리 사용된다. 여기서 APCER은 모델이 Fake 데이터를 Real 데이터로 잘못 분류한 비율이고, BPCER은 Real 데이터를 Fake 데이터로 잘못 분류한 비율을 말한다. 평균 오분류율은 Real과 Fake간 오류를 동일한 비중으로 고려하므로, 클래스 불균형 상황에서의 분류 모델의 전반적인 신뢰도를 정밀하게 평가할 수 있는 지표이다. 값이 낮을수록 모델의 전반적인 분류 정확도가 높은 것을 의미한다. 따라서 본 논문의 실험 결과에서 ACER가 낮은 모델을 더 뛰어난 성능을 가진 것으로 평가하였다. 위 지표들은 수식 (2), (3)과 같이 계산한다.

$$APCER = \frac{FN}{FN + TP}, \quad BPCER = \frac{FP}{FP + TN} \tag{2}$$

$$ACER = \frac{APCER + BPCER}{2} \tag{3}$$

4.2.2 AUC(Area Under the ROC Curve)

AUC는 ROC(Receiver Operating Characteristic) 곡선 아래 면적을 의미하며, 모델의 이진 분류 성능을 threshold에 관계 없이 종합적으로 평가할 수 있는 성능 지표이다. ROC 곡선은 TPR(True Positive Rate)와 FPR(False Positive Rate) 간의 관계를 나타내며, 1에 가까울수록 완벽한 분류에 가까움을 의미한다. 본 연구에서는 ACER을 주 지표로 사용하고, AUC를 보조 지표로 함께 활용하여 모델의 전반적인 탐지 성능을 다양한 지표를 통해 평가하였다. AUC는 수식 (4)와 같이 정의된다.

$$AUC = \int_0^1 TPR(x)\,dx \tag{4}$$

4.3 실험 세부 사항

본 연구의 실험은 629GiB의 메모리와 AMD EPYC 7642 CPU를 탑재한 서버 환경에서, NVIDIA A6000 GPU 한 대를 사용하여 수행되었다. 실험에 사용된 데이터는 전처리된 총 30,000장의 얼굴 이미지로, 이를 학습:검증:테스트 = 6:1:3의 비율로 분할하여 실험을 진행하였다. 학습 과정에서는 매 에폭(epoch)마다 검증 세트에 대해 모델의 성능을 평가하였으며, ACER 지표를 기준으로 가장 성능이 우수한 모델을 선택하여 테스트 세트에서 최종 성능을 평가하였다. 모델의 백본으로는 ViT-small_patch16_224 구조를 사용하였으며, 분류 손실 함수로는 Cross-Entropy Loss, 최적화 기법으로는 AdamW Optimizer를 적용하였다. 학습률은 ViT 모델에 대해 0.0001, RWM에 대해 0.00015로 설정하여 RWM이 더 빠르게 수렴하도록 하였다. 배치 크기는 128, 총 학습 에폭 수는 10으로 고정하여 모든 실험에서 일관된 조건을 유지하였다. 성능 평가는 앞서 설명한 바와 같이 ACER을 주요 지표로, AUC를 보조 지표로 활용하였다. 추가로, 모델의 일반화 성능을 평가하기 위한 실험을 진행할 때는 기존과 동일한 6,000개의 비디오로부터 프레임을 기존보다 두 배 많은 10개 지점에서 추출하여, 총 60,000장의 이미지를 구성하였다. 이는 보다 정확한 일반화 실험을 위해 더 넓은 데이터 분포를 반영하기 위함이며, 실제 환경에서의 적용 가능성을 검증하는 데 목적이 있다. 해당 실험에서는 과적합을 방지하고 안정적인 수렴을 유도하기 위해, 전체 학습률을 0.000001로 낮추어 설정하였다. 또한, DeepFakes, Face2Face, FaceShifter, FaceSwap, NeuralTextures의 5개 기법 각각에 대해 개별적으로 진행되었으며, 해당 기법의 데이터 전체와 real 데이터의 절반을 테스트 세트로 구성하였다. 나머지 real 데이터 절반과 나머지 4개 기법의 데이터를 학습용으로 사용하였고, 이를 다시 7:3 비율로 학습과 검증 세트로 분할하여 실험을 수행하였다. 이와 같은 설정은 특정 합성 기법을 학습하지 않은 상태에서 모델이 얼마나 새로운 기법에 대하여 잘 일반화되는지를 평가하기 위한 목적이다.

4.4 실험 결과

실험은 가중치 적용 방식에 따른 성능 비교, 5-point 및 Full-point 랜드마크 세트 활용 시 성능 차이 분석, 그리고 다양한 합성 기법에 대한 모델의 일반화 성능 평가, RWM과 CBAM의 평균 추론 속도 비교

들로 구성했다. 이를 통해 제안 기법의 유효성과 구성 요소별 성능 기여도를 정량적으로 검증하고자 하였다.

<Table 1>은 Full point 랜드마크 세트를 기준으로, Baseline과 RWM 적용 후 및 가중치 모드 별 Test 성능을 정리한 결과이다. 강조 부위를 제외한 나머지 패치에는 기본 가중치 1.0로 설정하였다. Baseline인 ViT−Small_Patch16_224 모델은 RWM이 적용되지 않은 상태에서 ACER 25.14%로 가장 높은 오류율을 기록하였다. 반면, 눈·코·입 부위에 고정 가중치된 가중치를 부여한 Fixed Mode에서는 가중치가 1.5배로 설정한 경우 ACER 22.50%, 1.2배로 설정한 경우 20.90%를 기록하여, 각각 Baseline 대비 2.64%, 4.24% 오류율 감소를 달성하였다. 이는 얼굴 주요 부위에 고정된 가중치를 부여하는 것만으로도 딥페이크 탐지 성능이 유의미하게 향상될 수 있음을 보여준다. 특히, 고정 가중치 방식 내에서도 1.2배 설정 시 1.5배보다 1.60% 낮은 ACER를 기록한 점은, 지나치게 높은 가중치보다는 적절한 수준의 가중치가 주요 랜드마크 뿐 아니라 주변 정보를 효과적으로 학습하도록 유도했음을 시사한다. 이러한 결과에 따라, 이후 모든 실험에서는 기본 가중치를 1.2배로 고정하여 실험을 진행하였다. 고정된 가중치가 아닌 학습 가능한 가중치를 사용하는 Single−Learnable 방식은 ACER가 18.06%로, Baseline 대비 7.08%, Fixed Mode 대비 2.84% 오류율이 낮아지는 성능 향상이 나타났다. 나아가, 눈·코·입 각각의 패치에 독립적인 가중치를 학습하는 Multiple Learnable 방식은 ACER 17.20%를 기록하며 Single 방식보다 약 0.86% 더 낮은 오류율을 기록하였다. 최종적으로, 본 논문의 제안 기법인 Multiple Learnable 가중치는 Baseline 대비 최대 7.94%의 ACER 개선을 이끌어냈으며, AUC 역시 87.91%에서 90.90%로 증가하는 성능 향상이 나타났다. 이러한 결과는 국소 왜곡에 대한 동적이고 정밀한 가중치 조정이 실제 환경에서도 딥페이크 탐지 성능을 유의미하게 증가시킬 수 있음을 입증했다.

〈Table 1〉 Performance comparison according to weighting methods based on full-point landmar set. The value in parentheses shows the change compared to baseline.

Weight Mode	ACER(%)	AUC(%)
Baseline	25.14	87.91
Fixed (x1.5)	22.50(−2.64%)	89.20
Fixed (x1.2)	20.90(−4.24%)	90.50
Single Learnable (x1.2)	18.06(−7.08%)	90.70
Multiple Learnable (x1.2)	17.20(−7.94%)	90.90

<Figure 5>는 가중치 모드 및 랜드마크 세트 구성 방식에 따라 입력 이미지에 적용된 패치 가중치 분포를 시각적으로 비교한 것이다. 동일한 색으로 표시된 패치들은 같은 가중치 값이 적용된 영역을 의미하며, 색이 다를 경우 서로 다른 독립적인 가중치가 부여된 영역임을 나타내도록 시각화를 구성하였다. (A)는 5−point 랜드마크 세트와 Single Weight 모드가 적용된 예시로, 눈동자 중심, 코끝, 입꼬리 등 다섯 개의 주요 지점에 해당하는 패치에만 동일한 가중치가 부여되어 있으며, 모두 동일한 색으로 표현되었다. (B)는 동일한 Single Weight 모드를 사용하지만 Full−point 세트를 적용한 경우로, 눈·코·입 전체 영역에 걸쳐 더 넓은 범위의 패치에 같은 가중치가 적용되었으며, 역시 하나의 색상으로 시각화되었다.

(C)는 Full−point 세트에 Multi Weight 모드를 적용한 예시로, 눈(노란색), 코(빨간색), 입(파란색) 각각의 부위에 서로 다른 가중치 값이 학습되어 독립적으로 적용되었음을 색상 차이를 통해 확인할 수 있다. 이러한 시각화를 통해 RWM이 실험 조건에 따라 의도한 대로 동작하고 있음을 정성적으로 확인할 수 있었다.

〈Figure 5〉 Visualization of patch emphasis by landmark set and weight mode
(A: 5−point set + Single Weight, B: Full−point set + Single Weight, C: Full−point set + Multi Weight)

<Table 2>는 양쪽 눈 동공, 코끝, 입술 양끝의 5개 대표 지점을 담은 5−point 랜드마크 세트와 부위의 모든 세부 지점을 포함한 Full−point 랜드마크 세트를 각각 사용하여 성능을 비교한 결과이다. Fixed 모드의 경우, 5−point 랜드마크 세트는 ACER 23.02%, AUC 90.9%를 기록한 반면, Full−point 세트는 ACER 20.90%, AUC 90.5%로 ACER가 2.12% 더 낮았다. Full−point 세트 기반의 Single−Learnable 모드 적용 시 ACER 18.06%, AUC 90.7%를 달성하여, 동일 모드에서의 5−point 세트 결과인 ACER 20.70%, AUC 89.9% 대비 ACER가 2.64% 향상되고 AUC도 0.8% 증가하였다. 마지막으로 Multi−Learnable 모드에서도, Full−point 세트는 ACER 17.20%, AUC 90.9%, 5−point 세트는 ACER 18.60%, AUC 90.1%로 각각 ACER 1.40%, AUC 0.8%의 성능 향상을 보였다. 이와 같이 동일한 가중치 모드에서도 Full−point 랜드마크 세트가 5−point 랜드마크 세트보다 더 낮은 ACER와 동등하거나 더 높은 AUC를 기록하는 것은, 세분화된 포인트들이 눈·코·입 부위의 미세한 합성 흔적을 더욱 정밀하게 포착하기 때문으로 해석할 수 있다.

〈Table 2〉 Performance comparison according to landmark set configuration.
The value in parentheses shows the change compared to baseline.

LandMark Set	Weight Mode	ACER(%)	AUC(%)
None	Baseline	25.14	87.91
5−point	Fixed(x1.2)	23.02(−2.12%)	90.90
	Single−Learnable(x1.2)	20.70(−4.44%)	89.90
	Multi−Learnable(x1.2)	18.60(−6.54%)	90.10
Full−point	Fixed(x1.2)	20.90(−4.24%)	90.50
	Single−Learnable(x1.2)	18.06(−7.08%)	90.70
	Multi−Learnable(x1.2)	17.20(−7.94%)	90.90

<Table 3>은 다양한 합성 기법에 대한 모델의 일반화 성능을 평가한 결과를 나타낸다. 본 실험에서는 FaceForensics++ 데이터셋에서 real 영상과 Deepfake, Face2Face, FaceShifter, FaceSwap, NeuralTextures 총 다섯 가지 딥페이크 기법을 사용해 4개 기법 및 Real 데이터 50%를 학습용으로, 남은 1개의 기법 및 절반의 Real 데이터를 테스트용으로 총 5개 기법을 번갈아 가며 평가했다.

HTER는 ACER와 동일한 계산식을 갖지만, 학습되지 않은 환경에서의 일반화 성능 평가에 자주 활용되는 지표로, 본 논문에서는 일반화 실험에서 ACER 대신 HTER(Half Classification Error Rate)라는 명칭을 사용하였다. 본 논문에서 제안한 RWM 방식은 모든 기법에서 Baseline 대비 HTER 감소를 달성하였다. DeepFakes의 경우 HTER는 18.64%로, Baseline 대비 3.21% 감소하였고, AUC는 93.83%로 소폭 향상되었다. Face2Face에서는 HTER가 24.09%로 1.08% 감소하였으며, FaceShifter와 NeuralTextures 역시 각각 36.56%, 33.83%로 각각 2.43%, 0.11%의 성능 개선을 보였다. FaceSwap에서는 다소 적은 폭인 0.15%의 다소 미약한 개선 효과를 보였다. FaceForensic++ 데이터셋에서 사용된 FaceSwap은 CNN 모델을 이용하여 소스 얼굴 전체를 잘라서 타겟 얼굴 위치에 덮어씌운 후, 피부 보정 및 블렌딩 처리를 적용하는 방식의 얼굴 교체 기법이다[15]. FaceSwap 계열 기법은 구현 구조가 비교적 단순하여, 합성 과정에서 눈·코·입과 같은 내부 안면 요소는 소스 얼굴의 형태가 그대로 보존되는 경우가 많다. 반면, 소스와 타겟의 피부 톤·윤곽선 차이를 보정하는 과정이 정교하지 못하기 때문에, 턱선 및 광대, 헤어라인 등 얼굴 외곽 경계부에서는 명확한 합성 흔적이 빈번하게 발생한다. <Figure 6>은 FaceForensics++ 데이터셋의 FaceSwap 기법이 적용된 이미지의 예시이다. 이 이미지는 얼굴 안쪽에 비해 이마, 관자놀이 근처에 부자연스러운 아티팩트가 눈에 띄게 존재하는 것을 볼 수 있다. 이러한 특성은 곧 본 연구에서 제안한 RWM의 한계와도 직결된다. RWM은 눈·코·입 영역의 미세한 변화를 강조하여 탐지 성능을 향상시키는 전략을 기반으로 하지만, FaceSwap 기법에서는 해당 부위가 원본과 유사하게 유지되므로 효과가 상대적으로 미약하게 나타난다. 이는 곧 딥페이크 기법마다 왜곡이 집중적으로 발생하는 부위가 상이하다는 사실을 보여준다.

결론적으로, Baseline과 본 논문의 제안 기법을 적용한 모든 Test 기법의 평균 HTER는 각 32.96%, 31.56%로 약 1.4%의 오류율이 감소한 성과를 보였다. AUC 평균 역시 약 0.12% 증가하며 소폭 향상된 결과를 보였다. 이는 본 논문의 제안 기법이 다양한 합성 유형에 대해 균형 있게 대응할 수 있음을 나타낸다. 하지만, 향후 탐지 모델은 눈·코·입과 같은 중심부뿐 아니라 턱선·이마·헤어라인 등 외곽 경계부를 포함한 다양한 영역을 기법에 맞게 선택적으로 강조할 수 있도록 후속 연구를 통해 발전시켜야 더욱 강건한 일반화 성능을 가질 수 있을 것이다.

〈Table 3〉 Generalization performance comparison by synthesis method.
The value in parentheses shows the change compared to baseline.

Test Method	MODE			
	Baseline		Multi−Learnable + Full−point set	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)
Deepfakes	21.85	93.78	18.64(−3.21%)	93.83
Face2Face	25.17	85.47	24.09(−1.08%)	85.57

FaceShifter	38.99	65.26	36.56(−2.43%)	65.43
FaceSwap	44.83	58.3	44.68(−0.15%)	58.4
NeuralTextures	33.94	71.98	33.83(−0.11%)	72.15
Average	32.96	74.96	31.56(−1.4%)	75.08(+0.12)

〈Figure 6〉 Example image with the FaceSwap method
(Left: Real; Right: FaceSwap-applied image)

Real **Fake**

<Table 4>는 CBAM과 RWM의 평균 추론 시간을 비교한 실험 결과를 정리한 표이다. 본 실험은 4.3절 실험 세부 사항에 기재된 GPU 자원 구성과 동일한 환경에서 수행되었으며, 비교 대상은 RWM을 적용한 ViT와 CBAM을 적용한 ViT이다. 224 x 224 리사이즈 처리된 동일한 단일 이미지를 대상으로 Warm−up 10회를 거친 뒤 100회 추론을 반복하여 평균 추론 처리 시간을 비교하였다. 측정은 이미지 텐서를 모델에 입력한 시점부터 출력이 반환될 때까지를 기준으로 하였고, GPU 환경에서의 정확한 시간 측정을 위해 추론 직후 연산 완료를 보장하도록 동기화하였다. RWM 측정의 경우 매 반복마다 캐시로 저장된 랜드마크를 조회하고 그 결과를 바탕으로 weight map을 생성·적용하는 과정까지 포함한 추론 시간을 기록하였으며, CBAM 측정은 랜드마크 조회가 전혀 포함되지 않은 순수한 추론 시간만을 기록하였다.

우선 랜드마크 캐시 조회 시간은 평균 약 0.0023 ms(≈2.3 μs)로 매우 적은 수치를 보였고 RWM 전체 추론 시간에 거의 영향을 미치지 않았다. 평균 추론 시간은 RWM이 12.01 ms, CBAM이 12.48 ms로 나타났다. RWM의 평균 추론 시간은 CBAM 대비 0.47ms(−3.91%) 더 낮았는데, 랜드마크 조회 시간과 weight map 생성·적용이라는 추가 연산을 포함하고도 RWM은 CBAM보다 오히려 더 낮은 평균 추론 시간을 달성했다. 이는 저장된 랜드마크 캐시의 조회가 마이크로초 단위로 매우 빠르게 동작하고 weight map 계산이 효율적으로 구현되어 제안 기법의 연산량 증가가 미미하며, 추론 속도 또한 CBAM보다 빠르다는 것을 입증했다. 따라서 RWM은 패치 어텐션 메커니즘으로 합성 흔적이 집중되는 패치를 선택적으로 강조하여 딥페이크의 탐지 민감도를 높이는 이점을 확보하면서도, 적은 연산량 및 빠른 추론 속도로 추후 실시간 또는 저지연 요구 환경에서의 적용 가능성을 보여주었다. 다만 본 비교는 단일 이미지·단일 배치 조건에서 수행된 결과이므로 입력 데이터 다양성 및 하드웨어 구성에 따라 절대적 수치는 달라질 수 있음을 명시한다.

〈Table 4〉 CBAM vs RWM Average inference time comparison (100 inferences).
The value in parentheses shows the change compared to baseline.

Backbone	Method	Mean inference time(ms)	Landmark lookup mean(ms)
ViT_small_patch 16_224	CBAM	12.48	N/A
	RWM	12.01(−3.91%)	0.0023

Ⅴ. 결 론

본 연구에서는 ViT 기반 딥페이크 탐지 모델의 국소 정보 학습 한계를 보완하기 위해, 얼굴의 눈·코·입 영역에 학습 가능한 가중치를 적용하는 RWM을 제안하였다. 이를 통해 모델은 딥페이크 생성 과정에서 주로 발생하는 얼굴 주요 부위의 미세한 합성 흔적을 보다 효과적으로 포착할 수 있도록 설계되었다. 실험은 FaceForensics++ 데이터셋을 기반으로, 가중치 모드(Fixed, Single−Learnable, Multi−Learnable), 랜드마크 세트(5−point vs. Full−point), 합성 기법별 일반화 성능 등 다양한 조건에서 진행되었다. 실험 결과, Multi−Learnable 모드와 Full−point 세트를 조합한 방식이 가장 우수한 성능을 나타냈으며, Baseline 대비 최대 7.94%의 ACER 개선과 AUC 향상을 달성하였다. 또한, 새로운 합성 기법에 대한 일반화 실험에서도 기존 모델보다 일관된 성능을 유지하며 RWM의 적용 효과를 입증하였다. 연산 비용 또한 기존 CBAM 방식보다 3.91% 감소한 평균 추론 시간을 보이며, 적은 연산량이 요구되는 실시간, 저지연 환경에서의 적용 가능성을 보여줬다. 본 논문의 제안 기법을 통하여 다양한 얼굴 합성 유형과 환경 변화에 강인하게 대응할 수 있는 딥페이크 탐지 모델 설계에 이바지할 수 있음을 기대한다.

Ⅵ. Acknowledgements

This research was supported by the 2025 Hanbat National University Academic and Cultural Research Foundation.

참 고 문 헌 (References)

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", Advances in Neural Information Processing Systems (Montreal,) 2014, pp. 2672-2680.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models", Advances in Neural Information Processing Systems (Vancouver,) 2020, pp. 6840-6851.

[3] Nan Jiang, Bangjie Sun, Terence Sim, and Jun Han, "Can I Hear Your Face? Pervasive Attack on Voice Authentication Systems with a Single Face Image," in Proceedings of the 33rd USENIX Security Symposium, (Philadelphia,), 2024, pp.1045-1062.

[4] F. Chollet, "Xception: Deep learning with depthwise separable convolutions", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Honolulu,) 2017, pp. 1251-1258.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Las Vegas,) 2016, pp. 770-778.

[6] 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 16×16 words: Transformers for image recognition at scale", International Conference on Learning Representations (Vienna,) 2021.

[7] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "Face X-ray for more general face forgery detection", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Seattle,) 2020, pp. 5000-5009.

[8] Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues", Proceedings of the European Conference on Computer Vision (Glasgow,) 2020, pp. 1-18.

[9] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module", European Conference on Computer Vision (Munich,) 2018, pp. 3-19.

[10] D. Nguyen, M. Astrid, E. Ghorbel, and D. Aouada, "FakeFormer: Efficient vulnerability-driven transformers for generalisable deepfake detection", Available: https://arxiv.org/pdf/2410.21964, 2025.09.10. confirmed.

[11] Z. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection", Information Fusion 64, Dec, 2020, pp. 131-148.

[12] H. Kim, J. Lee, Y. Jeong, H. Jang, and Y. J. Yoo, "Advancing cross-domain generalizability in face anti-spoofing: Insights, design, and metrics", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (Seattle,) 2024, pp. 970-979.

[13] F. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A framework for building perception pipelines", Available: https://arxiv.org/abs/1906.08172, 2025.09.10. confirmed.

[14] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to detect manipulated facial images", Proceedings of the IEEE/CVF International Conference on Computer Vision (Seoul,) 2019, pp. 1-11.

[15] I. Korshunova, W. Shi, J. Dambre, and L. Theis, "Fast face-swap using convolutional neural networks", Proceedings of the IEEE International Conference on Computer Vision (Venice,) 2017, pp. 3697-3705.

저 자 소 개

엄 기 원 (Ki-Won Eom)
준회원

2023년 3월~현재 : 국립한밭대학교 컴퓨터공학과 학사과정
관심분야 : 디지털 포렌식, 컴퓨터비전 등

김 동 수 (Dong-Su Kim)
준회원

2019년 3월~2025 2월 : 국립한밭대학교 컴퓨터공학과 학사
2025년 3월~현재 : 국립한밭대학교 소프트웨어융합 대학원 석사과정
관심분야 : 디지털 포렌식, 컴퓨터비전 등

오 태 우 (Tae-Woo Oh)
평생회원

2007년 2월 : 아주대학교 정보및컴퓨터공학부 학사
2009년 2월 : 한국과학기술원 전산학과 석사
2012년 2월 : 한국과학기술원 전산학과 박사
2011년~현재 : 국가보안기술연구소 책임연구원
관심분야 : 멀티미디어, 콘텐츠 보안, 디지털 워터마킹, 디지털 포렌식 등

장 한 얼 (Haneol Jang)
준회원
2012년 2월 : 아주대학교 정보컴퓨터공학부 학사
2018년 2월 : 한국과학기술원 전산학부 박사
2018년 4월~2018년 10월 : 네이버 클로바 인공지능 연구원
2018년 10월~2020년 8월 : 국가보안기술연구소 선임연구원
2020년 9월~현재 : 국립한밭대학교 컴퓨터공학과 교수
관심분야 : 디지털포렌식, 컴퓨터비전, 머신러닝 등