논문 22-47-03-16 The Journal of Korean Institute of Communications and Information Sciences '22-03 Vol.47 No.03
https://doi.org/10.7840/kics.2022.47.3.551

샴 네트워크와 동형암호를 이용한 개인정보 보호 얼굴 검증 시스템

조 현 진•, 강 효 은*, 심 준 석*, 홍 윤 영*, 김 호 원°

Privacy Protection Face Verification System Using Homomorphic Encryption and Siamese Network

Hyun-jin Cho•, Hyo-eun Kang*, Jun-seok Sim*, Yoon-young Hong*, Ho-won Kim°

요 약

이미지 인식에서 좋은 성능을 보이고 있는 딥러닝은 최근 다양한 분야에서 필수적인 요소로 작용하고 있으며, 생체인식의 중요성이 증가함에 따라 얼굴 검증에 적합한 딥러닝 모델들이 연구되고 있다. 그러나 딥러닝 모델의 학습에는 사용자의 얼굴 이미지를 그대로 사용한다. 이 경우 딥러닝 모델의 취약점인 Inversion attack에 의해 사용자의 얼굴 이미지가 유출이 될 가능성이 존재하며 프라이버시가 보호되지 않는 점을 암시한다. 따라서 사용자의 개인정보인 얼굴 이미지의 프라이버시를 보호하면서 딥러닝 모델의 정상적인 검증을 가능하게 하는 기술이 필요하다. 본 논문에서는 사용자의 프라이버시를 보호하기 위하여 딥러닝 학습 및 추론에 사용되는 얼굴 이미지에 Paillier 동형암호화를 이용한 암호화를 적용한다. 또한 특징 벡터 간의 유사도를 이용하여 추론을 하는 샴 네트워크를 사용함으로써 구분하기 어려운 암호화 데이터에 대해서도 학습 및 검증 과정에서의 높은 정확도를 보이도록 하였다.

Key Words : Siamese Network, Homomorphic Encryption, Privacy Invasion, Face Verification, Deep Learning

ABSTRACT

Deep learning, which has shown good performance in image recognition, has recently become an essential element in various fields, and as the importance of biometric recognition increases, deep learning models suitable for face verification are being studied. However, deep learning model uses the user's face image to learn. In this case, it implies that the Inversion attack, the vulnerability of the deep learning model may extract the user's face image and not protect privacy. Therefore, there is a need for a technology that enables normal verification of the deep learning model while protecting the privacy of the face image, which is the user's personal information. In this paper, to protect users' privacy, encryption using Paillier isomorphic encryption is applied to facial images used for deep learning and reasoning. In addition, by using a Siamese network that infers using similarities between feature vectors, high accuracy in the learning and verification process was also shown for encrypted data that is difficult to distinguish.

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획 평가원의 지원을 받아 수행된 연구임 (2019-0-01343, 융합보안핵심인재양성사업).
• First Author : Pusan National University, hyunjin@islab.re.kr, 학생회원
° Corresponding Author : Pusan National University, howonkim@gmail.com, 종신회원
* Pusan National University, hyoeun@islab.re.kr, 학생회원; junseok@islab.re.kr, yoonyoung@islab.re.kr
논문번호 : 202110-269-0-SE, Received September 29, 2021; Revised November 17, 2021; Accepted December 24, 2021

Ⅰ. 서 론

얼굴 검증(Face verification)은 컴퓨터 비전의 핵심 문제 중 하나로 활발히 연구되어 왔다[1]. 얼굴 검증의 목표는 두 개의 얼굴 영상 또는 이미지가 주어지면 동일한 얼굴인지 확인하는 것이다. 최근 딥러닝의 기술 발달로 얼굴 검증과 딥러닝을 접목시킨 다양한 연구[2-5]들이 제시되고 있다. 딥러닝을 이용한 얼굴 검증 기술은 스마트폰의 FaceID, 모바일 앱 로그인 및 결제 확인[6] 등 보안 시스템에 적용되고 있다. 그러나 Goodfellow 등[7]에서 보여준 적대적 예제(Adversarial example)를 시작으로 보안 취약점들이 지속적으로 제시되고 있다. 딥러닝 기술의 보안 취약점을 이용한 Inversion attack 공격 기법[8-10]이 존재한다. Inversion attack은 딥러닝 모델에 반복적으로 추론 요청을 보내고 추론 결과를 합쳐서 학습에 사용된 데이터를 복원하는 기법이다. 공격자가 Inversion attack을 사용자 인증 시스템에 시도할 경우 개인 정보 탈취 위험성이 존재한다. 또한 딥러닝 서비스는 퍼블릭 클라우드를 활용하여 제공되기도 하는데, 이 때 네트워크 전송 과정에서 데이터를 도청하는 스니핑(Sniffing) 공격 등의 위험성도 존재한다. 딥러닝 모델에 대한 보안 취약점을 이용한 공격으로부터 데이터의 프라이버시 보호를 위해 Munirah 등[11]은 인공위성으로 촬영한 이미지에 동형암호를 적용한 모델 학습 기법을 제안했다. 본 연구에서 제안하는 시스템은 동형암호를 이용한 암호화를 통해 얼굴 이미지 데이터의 프라이버시를 보호한다. 또한 소량의 데이터를 사용하여 학습하여도 높은 성능을 보일 수 있는 One-shot learning 기반 샴 네트워크(Siamese network)를 사용한다. 샴 네트워크는 검증 과정에서 특징 벡터(Feature vector) 간의 차이를 이용하므로 검증하기 어려운 암호화한 데이터에 대해서도 기존 딥러닝 모델보다 높은 검증 정확도를 보인다. 본 논문의 구성은 2장에서 본 논문과 관련된 연구들에 대해 설명하고 3장에서 제안하는 기술의 전체 구조를, 4장에서 본 논문에서 제안하는 기술의 성능을 검증하기 위한 실험을 설명하며, 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 본 론

2.1 관련 연구

2.1.1 동형암호화를 적용한 얼굴 검증 시스템

Ma[5] 등은 얼굴의 특징을 통한 등록 및 검증 시스템인 DeepID를 제안하였다. DeepID 시스템은 CNN(Convolutional Neural Network) 기반 DeepID network와 암호화 모듈로 구성된다. 암호화 모듈은 DeepID network를 통해 추출한 얼굴 특징을 이진화 및 Pailler 암호화를 진행하여 DeepID 시스템에 저장한다. 저장한 암호화된 얼굴 특징들은 DeepID 시스템의 검증 과정에서 사용한다. DeepID 시스템은 사용자가 검증 받고자 하는 얼굴 이미지를 입력하면, DeepID network와 암호화 모듈을 거쳐 암호화된 특징을 추출한다. 시스템은 저장된 특징들과 출력된 특징간의 해밍 거리(Hamming distance)기반 유사도 값이 임계값 이하일 경우 동일 인물이라고 판별한다. 저자가 제안한 시스템은 약 30,000개의 얼굴 이미지로 네트워크를 학습하며 사용자가 얼굴 이미지를 입력하는 과정에서 스니핑 공격으로 인해 이미지 탈취 위험성이 존재한다.

2.1.2 완전동형암호를 이용한 얼굴 검증 시스템

Naresh 등[12]은 얼굴 이미지의 특징에 대하여 완전동형암호(Fully homomorphic encryption)를 통하여 암호화된 이미지로 얼굴 인식이 가능한 시스템을 제안했다. 시스템은 Fan-Vercauteren 완전동형암호화[13]와 딥러닝 얼굴 인식 모델 Facenet[14]을 사용하여 등록 및 인식을 수행한다. 등록 과정에서 사용자는 비밀키-공개키 쌍을 발급받고 공개키를 사용해서 Facenet에서 추출한 특징을 암호화한다. 암호화된 특징은 아이디, 공개키와 같이 전송되어 등록된다. 사용자가 얼굴 매칭을 수행하기 위해서는 등록 과정과 동일하게 특징 추출 및 암호화를 진행하여 나온 출력 값을 아이디와 함께 시스템에 전송한다. 시스템은 등록된 특징들과 입력된 특징 간의 코사인 유사도(Cosine similarity) 기반 비유사성 점수를 계산한다. 시스템은 계산된 비유사성 점수를 요청된 ID의 공개키로 암호화하여 사용자에게 전송한다. 사용자는 자신의 비밀키로 비유사성 점수를 복호화하여 자신과 데이터베이스 내에 존재하는 ID 간의 얼굴 인식 결과를 확인한다.

2.1.3 CNN을 이용한 얼굴 검증 시스템

Cheng 등[15]은 얼굴 검증 시 대상 인물의 포즈, 노화 정도, 화장품 사용 유무 또는 조명의 변화에 따라 정확도가 감소하는 점을 해결하는 알고리즘을 제안하였다. 제안된 알고리즘에서 사용한 데이터 세트는 이미지를 검증하는데에 있어서 비교적 쉽고 개수가 적은 CASIA-WebFace[16]와 상대적으로 어려운 IJB-A[17]이다. CASIA-WebFace 와 IJB-A 데이터 세

트에 대해 각 이미지 데이터의 얼굴을 7개의 특징 포인트(왼쪽 눈 모서리 2개, 오른쪽 눈 모서리 2개, 코 끝, 입 모서리 2개)를 이용해서 영역별로 정렬한다. CASIA-WebFace를 이용하여 CNN 모델을 학습시키고, IJB-A의 학습데이터와 CNN의 Joint Bayesian Metric을 구한 뒤 학습시킨다. 한 쌍의 테스트 이미지 세트가 주어지면, 앞서 학습된 Joint Bayesian Metric과 CNN을 이용해서 유사성 점수를 계산한다. 실험 결과는 IJB-A의 테스트 데이터 세트를 이용한 정확도 측정 시, 제안된 CNN의 성능이 FV(Fisher vector)를 기반으로 하는 얼굴 검증 시스템 및 기타 상용화된 얼굴 검증기보다 우수한 성능을 보여준다.

2.2 시스템 구조

2.2.1 Paillier 암호화

그림 1은 원본 이미지와 Paillier 암호[18]를 이용한 동형암호화를 적용한 이미지이다. 본 시스템에서는 사용자의 프라이버시를 보호하기 위해 비대칭 암호화 알고리즘인 Paillier 암호화 알고리즘을 사용하여 얼굴 이미지를 암호화한다. Paillier 암호체계는 임의의 양의 정수 p보다 작은 양의 값의 정수 a와 b에 대해서 $a^x \equiv b(\mathrm{mod}\, p)$를 만족하는 x를 구하는 이산 로그 문제(Discrete logarithm problem)의 원리를 이용하여 설계된 암호이다. 이산 로그 문제는 NP(Non-deterministic Polynomial)문제로 공격자가 a, b, p의 정보를 알고 있어도 x를 다항시간에 계산할 수 없다. Paillier 암호체계는 키 생성, 암호화, 복호화의 3단계를 가지며 표 1, 표 2, 표 3과 같다. 표 1의 키 생성 알고리즘을 통해 생성된 비밀키는 공격자가 공개키 (n, g)를 알고 있어도 이산로그문제의 계산복잡성 때문에 쉽게 구할 수 없다. Paillier 암호체계의 특징은 비밀키에 대한 정보가 없어 복호화가 불가능하여도 암호문 c의 상수곱과 암호문 간의 덧셈 결과

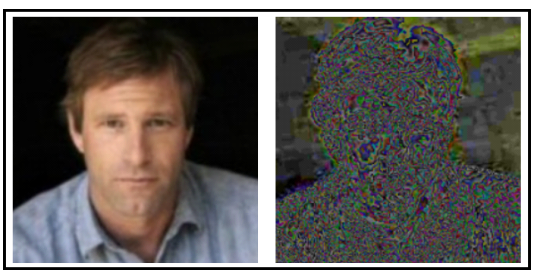


그림 1. 원본 이미지(왼쪽), 암호화된 이미지(오른쪽)
Fig. 1. Original Image(left), Encrypted Image(Right)

표 1. 키 생성 알고리즘
Table 1. Key Generation algorithm

Algorithm 1. Key Generation
1. *Choose* p and q *where* $\gcd(pq, p-1, q-1) = 1$ 2. $n = p \times q$, $\lambda = lcm(p-1, q-1)$ 3. *Select a integer randomly as* g *where* $g \in Z^* n^2$ 4. *Define the function*: $L(x) = (x-1)/n$ 5. *Verify the existence of the following modular μltiplicative inverse to ensure to ensure that* n *divedes* $g's$ *order*: $\mu = L(g^\lambda \bmod (n^2))^{-1} \bmod (n)$ *The public key used*: (n, g) *The private key used*: (λ, μ)

표 2. 암호화 알고리즘
Table 2. Encryption algorithm

Algorithm 2. Encryption
1. *Let* m *be a messge to be encrypted where* $0 \le m < n$ 2. *Select at random an integer* r *where* $0 < r < n$ and $r \in Z^* n^2$ 3. *Compute cipher text as*: $c = g^m r^n mod\ n^2$

표 3. 복호화 알고리즘
Table 3. Decryption algorithm

Algorithm 3. Decryption
1. *Let* c *be the cipher text to decrypt, where* $c \in Z^* n^2$ 2. *Compute the plaintext message as*: $m = L(c^\lambda \bmod n^2)$.

가 평문 m의 연산 결과와 동일한 동형연산이 가능한 부분동형암호(Partial homomorphic encryption)라는 점이다.

$$c^a = (g^m \times r^n)^a = (g^{ma} \times r^{an}) \bmod n^2 \quad (1)$$

식 1에 암호문 c_1과 상수 a를 대입하여 상수곱 ma를 구할 수 있다.

$$\begin{aligned} c_1 \times c_2 &= (g^{m1} \times r_1^n \times g^{m2} \times r_2^n) \bmod n^2 \\ &= (g^{m1+m2} \times (r_1 \times r_2)^n) \bmod n^2 \end{aligned} \quad (2)$$

또한 식 2에 암호문 c_1와 c_2를 대입하여 평문 m_1과 m_2의 덧셈 결과인 $m_1 + m_2$를 구할 수 있다.

본 논문에서는 Paillier 암호화에 사용되는 소수 p와 q의 길이를 10으로 설정하여 $n = p \times q$의 계산에 소요되는 시간을 최소화한다. $\gcd(pq, p-1, q-1) = 1$을 만족하는 임의의 2개 소수 p와 q를 입력으로 하여 표 1의 키생성 알고리즘을

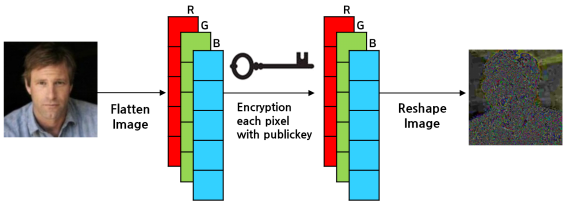


그림 2. 이미지 암호화의 과정
Fig. 2. Process of image encryption

통해 공개키와 비밀키를 생성한다. 키 생성 시 사용되는 g는 n^2의 계산에 소요되는 시간을 최소화하기 위하여 $p \times q+1$로 설정한다. 표 2의 알고리즘을 적용하기 위해 이미지를 1차원 배열로 변형한 후 공개키와 임의의 값을 이용하여 암호화한다. 암호화된 1차원 형태의 배열은 모델에 입력하기 위해 이미지의 원래 모양으로 변형하여 이미지의 암호화를 완료한다. 전체 이미지 암호화 과정은 그림 2와 같다.

2.2.2 샴 네트워크

Gregory 등[19]은 하나의 데이터로만 학습하여도 정확한 추론이 가능한 One-shot learning 기반의 샴 네트워크를 제안하였다. 샴 네트워크는 동일한 파라미터를 공유하는 트윈 네트워크(Twin network)를 포함하는 딥러닝 모델이다. 그림 3과 같이 트윈 네트워크 중 상위 네트워크는 입력으로 검증받고자 하는 이미지(Input Image)가 주어지고, 하위 네트워크는 입력으로 저장되어있는 검증용 참조 이미지(Reference Image)들이 주어진다. 트윈 네트워크는 두 이미지로부터 특징 벡터(Feature vector)를 추출하여 Distance layer로 전달한다. Distance layer는 2개의 특징 벡터간의 L_1 거리를 계산하여 벡터 간의 유사도를 계산하는 Prediction layer으로 전달한다. Prediction layer에서 특징 벡터 간의 유사도를 구하는 식은 식 3과 같다.

$$\sum_{j}\left(\alpha_j\left|h_1^{(j)}-h_2^{(j)}\right|\right) \tag{3}$$

h_1과 h_2는 상위 네트워크와 하위 네트워크에서 출력된 각각의 특징 벡터, α_j는 특징 벡터의 j번째 특징 간 L_1거리에 곱해지는 Prediction layer의 파라미터이다. Prediction layer는 특징 벡터 간의 유사도에 시그모이드 함수를 적용하여 0과 1사이의 값을 출력하며 이는 이미지와 참조 이미지의 클래스가 동일할 확률이다. 그림 3과 같은 구조로 입력 이미지가 참조 이미지와 동일한 클래스이면 특징 벡터 간의 유사도가 커지도록, 참조 이미지와 동일하지 않은 클래스이면 특징 벡터 간의 유사도가 작아지도록 학습을 하여 새로운 데이터에 대해서 재학습 없이 높은 성능의 얼굴 검증이 가능하다.

2.2.3 시나리오

제안하는 시스템은 등록된 사람들의 얼굴 이미지들과 사용자의 얼굴 이미지 간의 비교를 통해 각각의 확률들을 출력하고, 그 중 설정된 임계값을 넘는 확률이 있을 경우 시스템에 등록된 사람이라고 판단한다. 프라이버시 보호를 고려하지 않은 시스템의 경우, 아래 그림 4와 같이 Inversion attack으로 인해 개인정보가 유출될 수 있다.제안하는 시스템에서는 딥러닝 모델의

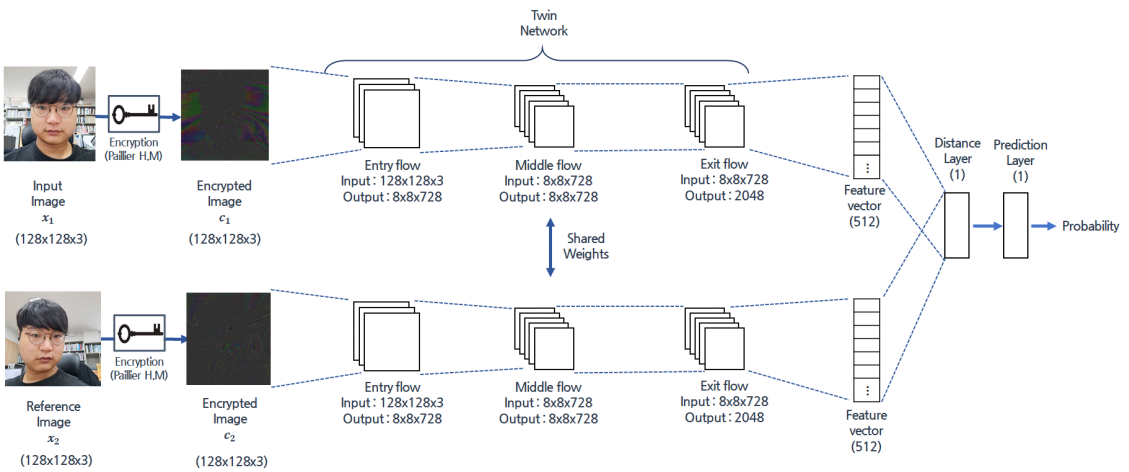


그림 3. 시스템 구조
Fig. 3. System architecture

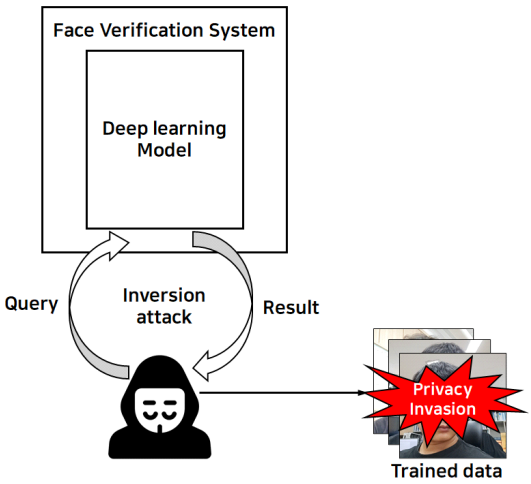


그림 4. 프라이버시 침해 시나리오
Fig. 4. Privacy invasion scenario

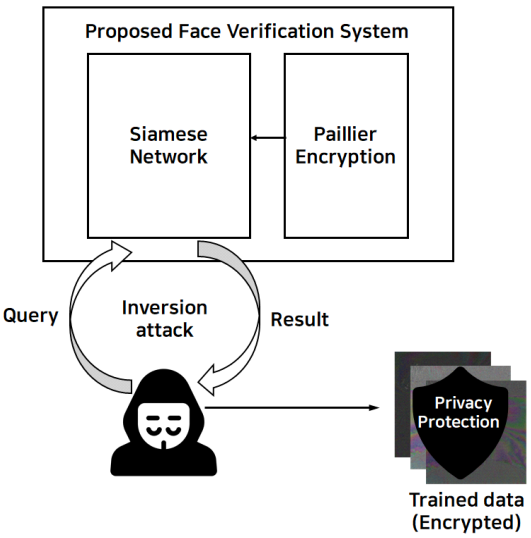


그림 5. 프라이버시 보호 시나리오
Fig. 5. Privacy protection scenario

학습 및 추론 시 이미지에 암호화를 적용하므로, 그림 5와 같이 Inversion attack이 발생하여도 암호화된 데이터가 유출된다. 공격자는 비밀키에 대한 정보가 없으므로 데이터의 복호화를 진행할 수 없으며 프라이버시가 보호된다.

Ⅲ. 실 험

3.1 모델 구조

본 논문에서 제안하는 얼굴 검증 시스템의 샴 네트워크 모델은 $128 \times 128 \times 3$ 크기의 이미지 쌍을 입력받는 트윈 네트워크로 구성된다. 이미지 쌍은 검증하고자 하는 입력 이미지와 얼굴 검증을 위해 사용될 참조 이미지로 구성된다. 2개의 이미지는 트윈 네트워크에 입력되기 전 동형암호화를 거친다. 본 논문에서 인셉션 모듈(Inception module)[20]을 사용하는 XceptionNet[21]과 MobileNet[22]을 트윈 네트워크로 사용하는 2개의 샴 네트워크 모델을 사용한다. 본 논문에 사용된 XceptionNet과 MobileNet은 ImageNet 데이터 세트[23]로 사전학습된 가중치를 사용한다. 시그모이드 출력은 입력 이미지 쌍이 동일 인물의 얼굴인지 혹은 서로 다른 인물의 얼굴인지를 판단하는 확률로 사용된다. 사용자에 의해 정해진 임계값을 기준으로 확률이 낮으면 다른 인물, 높으면 같은 인물로 판단한다.

3.2 실험 환경

본 연구에서는 얼굴 검증 모델 학습을 위해 Intel Core i9-10900k 3.70GHZ CPU, 메모리 32GB, Geforce RTX 3090 GPU를 활용하였다. Tensorflow 라이브러리 및 기계 학습 오픈 소스 라이브러리인 Scikit-learn를 기반으로 구현 및 검증하였다.

3.3 모델 학습

본 연구에서는 얼굴 검증 모델의 학습을 위해 LFW(Labeled Faces in the Wild) 데이터 세트[24]를 사용한다. LFW 데이터 세트는 5,749명의 사람에 대해 13,233개의 이미지로 구성된다. 학습 및 검증을 위해 1,680명에 대해 동일 인물 및 동일하지 않은 인물로 구성된 이미지 쌍 데이터 세트을 구성했다. 동일 인물로 구성된 이미지 쌍에는 1, 동일하지 않은 인물로 구성된 이미지 쌍에는 0의 라벨을 부여했다. 학습 시 Adam을 Optimizer로써 사용했으며, Binary cross entropy를 손실함수(Loss function)로 사용했다.

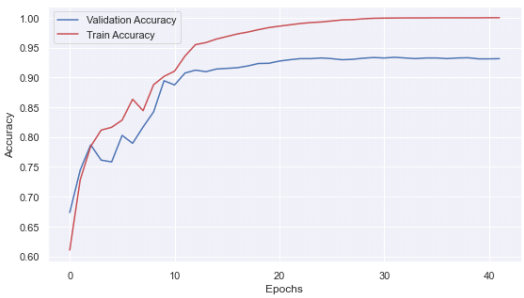


그림 6. 훈련 및 검증 정확도
Fig. 6. Training and validation accuracy

3.4 실험 결과

본 논문에서는 암호화 이미지의 얼굴 검증 성능의 평가를 위해 원본 이미지와 암호화 이미지의 얼굴 검증 결과의 ROC curve (Receiver Operating Characteristic curve)를 사용한다. 그림 7은 XceptionNet을 트윈 네트워크로 한 모델의 원본 이미지와 암호화 이미지에 대한 얼굴 검증 결과의 ROC curve 간의 비교이다. 원본 이미지와 암호화 이미지의 AUC(Area Under the ROC curve)는 각각 0.99, 0.97 이다. Park[25]에 따르면 AUC는 $0.5 < AUC \leq 0.7$ 일 때 낮은 정확도(less), $0.7 < AUC \leq 0.9$ 일 때 중등도(moderate), $0.9 < AUC \leq 1.0$ 일 때 높은 정확도(high)를 가진다. 따라서 원본 이미지에 대한 얼굴 검증 결과와 암호화 이미지에 대한 얼굴 검증 결과는 높은 정확도를 가진다.

표 4는 트윈 네트워크를 구성하는 CNN 모델 및 모델 출력의 임계값에 따른 정확도를 나타낸다. XceptionNet은 암호화 이미지에서 0.88~0.93의 정확도를, MobileNet의 경우, 암호화 이미지에서 0.88~0.9의 정확도를 보였다. 두 모델의 정확도를 비교했을 때, XceptionNet을 트윈 네트워크로 사용한 샴 네트워크가 더 정확한 얼굴 검증 결과를 보였다. 또한 XceptionNet을 트윈 네트워크로 사용할 경우, 임계값에 관계 없이 암호화 이미지에 대해서도 높은 정확도의 얼굴 검증결과를 확인할 수 있었다.

그림 6은 사전학습된 XceptionNet을 트윈 네트워크로 한 모델의 학습 및 검증 정확도를 나타낸다. 학습 및 검증 데이터 세트는 전체 이미지 쌍 데이터 세트를 7:3의 비율로 구성했다. 임계값을 0.7로 설정했을 때, 학습 데이터 세트의 최종 정확도는 99.4%, 검증 데이터 세트의 최종 정확도는 93.03%을 달성했다.

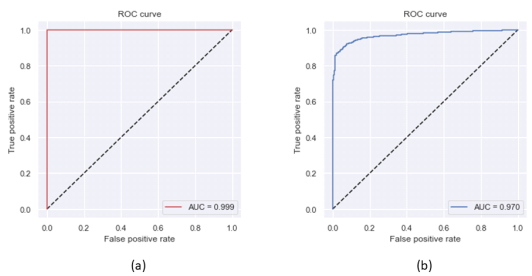


그림 7. ROC curve 비교, (a) 원본 이미지를 검증한 경우의 ROC curve, (b) 암호화된 이미지를 검증한 경우의 ROC curve
Fig. 7. ROC curve comparison, (a) ROC curve of validation with original images, (b) ROC curve of validation with encrypted images

표 4. 트윈 네트워크와 임계값에 따른 모델의 정확도 비교
Table 4. Comparison of model accuracy according to twin networks and thresholds

Threshold	XceptionNet	MobileNet	XceptionNet (encrytion)	MobileNet (encrytion)
0.5	0.99	0.99	0.88	0.87
0.6	0.99	0.99	0.90	0.88
0.7	0.99	0.99	0.92	0.90
0.8	0.99	0.99	0.93	0.90
0.9	0.99	0.96	0.92	0.89

Ⅳ. 결 론

본 논문에서는 사람의 얼굴 이미지에 대한 프라이버시를 보호하는 얼굴 검증 시스템을 제안하였다. 제안된 기술은 Paillier 동형암호를 사용하여 암호화된 이미지만을 학습 및 검증에 사용함으로써 딥러닝 모델 공격으로 인한 데이터 유출되는 이미지들에 대하여 프라이버시를 보호한다. 또한 One-shot learning 및 특징 벡터 간 유사도 기반으로 추론하는 샴 네트워크를 사용함으로써 소량의 암호화된 얼굴 이미지 데이터로 학습하여도 90% 이상의 뛰어난 검증 정확도를 보이는 점을 검증하였다. 향후, 제안한 시스템에서는 부분동형암호인 Paillier 암호 대신 검증 시스템에 더 적합한 동형암호를 이용한 암호화를 진행할 예정이다.

References

[1] A. K. Sao and B. Yegnanarayana, "Face verification using template matching," *IEEE Trans. Inf. Forensics and Secur.*, vol. 2, no. 3, pp. 636-641, 2007.

[2] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," in *Proc. IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 38, no. 10, Oct. 2016.

[3] A. Lebedev, et al., "Face verification based on convolutional neural network and deep learning," *2017 IEEE EWDTS*, 2017.

[4] S. Chen, et al., "Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices," *Chin. Conf. Biometric Recognition*, Springer, Cham, 2018.

[5] Y. Ma, et al., "A secure face-verification scheme based on homomorphic encryption and

deep neural networks," *IEEE Access*, vol. 5, pp. 16532-16538, 2017.

[6] J. Park, "인공지능과 얼굴 정보 처리 기술" *ITFIND*, vol. 1989, pp. 17-28, 2021.

[7] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[8] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Comput. and Commun. Secur.*, pp. 1322-1333, Oct. 2015.

[9] R. Shokri, et al., "Membership inference attacks against machine learning models," *2017 IEEE Symp. Secur. and Privacy*, 2017.

[10] N. Carlini, et al., "Extracting training data from large language models," *USENIX Security 21,* 2021.

[11] M. Alkhelaiwi, et al., "An efficient approach based on privacy-preserving deep learning for satellite image classification," *Remote Sensing*, vol. 13, no. 11, 2021.

[12] V. N. Boddeti, "Secure face matching using fully homomorphic encryption," *2018 IEEE 9th Int. Conf. BTAS*, 2018.

[13] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic encryption," *IACR Cryptology ePrint Archive*, vol. 2012, pp. 144, 2012.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vision And Pattern Recognition,* 2015.

[15] J.-C. Chen, V. M. Patel, and R. Chellappa, "Unconstrained face verification using deep CNN features," *2016 IEEE WACV*, 2016.

[16] D. Yi, et al., "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[17] B. F. Klare, et al., "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark A," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition,* 2015.

[18] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," *EUROCRYPT '99*, pp. 223-238, Springer, Berlin, Heidelberg, Apr. 1999.

[19] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," *ICML Deep Learn. Wkshps.*, vol. 2, Lille, France, 2015.

[20] C. Szegedy, et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2015.

[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2017.

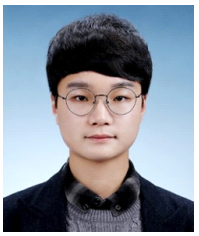
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in NIPS*, 25, pp. 1097-1105, 2012.

[23] A. G. Howard, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv: 1704.04861*, 2017.

[24] G. B. Huang and E. Learned-Miller, "*Labeled faces in the wild: Updates and new reporting procedures*," Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. 14.003, 2014.

[25] S.-I. Pak and T.-H. Oh, "Application of receiver operating characteristic (ROC) curve for evaluation of diagnostic test performance," *J. Veterinary Clinics*, vol. 33, no. 2, pp. 97-101, 2016.

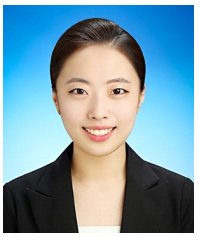
조 현 진 (Hyun-jin Cho)



2020년 : 부산대학교 정보컴퓨터공학과 졸업
2020년~현재 : 부산대학교 정보융합공학과 석사과정
<관심분야> 인공지능, 컴퓨터 비전, 적대적 공격

[ORCID:0000-0003-1728-1702]

강 효 은 (Hyo-eun Kang)



2017년 : 부산대학교 IT 응용공학과 졸업
2018년~현재 : 부산대학교 정보융합공학과 석박통합과정
<관심분야> 인공지능, 컴퓨터 비전, 산업용 AI

[ORCID:0000-0002-9651-7439]

심 준 석 (Jun-seok Sim)

2020년 : 해양대학교 제어계측공학과 졸업
2021년~현재 : 부산대학교 정보융합공학과 석사과정
<관심분야> 인공지능, 자연어 처리, 컴퓨터 비전

[ORCID:0000-0001-5762-3362]

홍 윤 영 (Yoon-yeong Hong)

2016년~현재 : 부산대학교 정보컴퓨터공학과 재학 중
<관심분야> 인공지능, 컴퓨터 비전, 산업용 AI

[ORCID:0000-0001-6045-8505]

김 호 원 (Ho-won Kim)

1993년 : 경북대학교 전자공학과 졸업 (학사)
1995년 : 포항공과대학교 전자전기공학과 졸업 (석사)
1999년 : 포항공과대학교 전자전기공학과 졸업 (박사)
1998~2008년 : 한국전자통신연구원 (선임연구원/팀장)
2008년~현재 : 부산대학교 정보컴퓨터공학부 교수
2014년~현재 : 부산대학교 사물인터넷연구센터 센터장
2020년~현재 : 부산대학교 지능형융합보안대학원 책임교수
2020년~현재 : 부산대학교 블록체인 플랫폼 연구센터 센터장
<관심분야> 인공지능, 정보보호, 블록체인 IoT 등
[ORCID:0000-0001-1728-1702]