정 보 보 호 학 회 지
제 29 권 제3호, 2019. 6

딥러닝 기반 얼굴인식 모델에 대한 변조 영역 제한 기만공격

류 권 상*, 박 호 성**, 최 대 선***

요 약

최근 딥러닝 기술은 다양한 분야에서 놀라운 성능을 보여주고 있어 많은 서비스에 적용되고 있다. 얼굴인식 또한 딥러닝 기술을 접목하여 높은 수준으로 얼굴인식이 가능해졌다. 하지만 딥러닝 기술은 원본 이미지를 최소한으로 변조시켜 딥러닝 모델의 오인식을 발생시키는 적대적 예제에 취약하다. 이에 따라, 본 논문에서는 딥러닝 기반 얼굴인식 시스템에 대해 적대적 예제를 이용하여 기만공격 실험을 수행하였으며 실제 얼굴에 분장할 수 있는 영역을 고려하여 설정된 변조 영역에 따른 기만공격 성능을 분석한다.

Ⅰ. 서 론

얼굴인식 기술은 입력 얼굴 영상이 어떤 인물인지 판별하는 기술로 출입국 심사, 결제 시스템, 단말 잠금 해제 등과 같은 실제 서비스에 적용되고 있다. 실제 환경에서 수집되는 얼굴 영상은 다양한 표정 및 조명 변화, 원거리 촬영, 해상도, 블러(blur) 등으로 인해 얼굴인식 성능이 떨어진다는 문제점을 가지고 있다. 이러한 문제들은 딥러닝 기술로 대용량의 데이터를 학습시켜 해결되었다. 딥러닝 기반의 얼굴인식 기술은 다양한 데이터 환경에서도 높은 성능의 얼굴인식이 가능하게 되었으며, 사람의 인지 수준을 넘어서는 연구 또한 늘어나고 있다. [1,2,3,4,5,6,7]

최근 딥러닝 기술은 음성 인식 [8], 침입탐지 [9] 등과 같은 다양한 분야에서 놀라운 성능을 보여주고 있어 많은 관심을 받고 있다. 따라서 딥러닝 기술을 이용한 인공지능 서비스가 증가함과 동시에 딥러닝 기술에 대한 보안 문제 또한 관심을 받고 있다.

이러한 딥러닝 기술에 대한 보안 문제로 적대적 공격(adversarial attack)이 있다. [10] 적대적 공격은 입력 데이터를 최소한으로 변조시켜 딥러닝 모델이 원래 class가 아닌 다른 class로 오인식하게 만드는 적대적 예제(adversarial examples)를 생성하여 공격하는 기법이다. 예를 들어, 자율주행차량에서 딥러닝 기술을 사용한다면 적대적 예제는 자율주행차량이 정지 표지판을 속도제한 표지판으로 오인식하여 인명피해를 발생시킬 수 있다. [11]

본 논문에서는 적대적 공격을 통해 생성된 적대적 예제를 이용하여 딥러닝 기반 얼굴인식 모델에 대해 기만공격을 수행하고 실제 얼굴에 분장할 수 있는 영역을 고려하여 설정된 변조 영역에 따른 기만공격의 성능을 분석한다.

본 논문의 구조는 다음과 같다. 2장에서는 적대적 예제를 설명하고 3장에서는 딥러닝 기반 얼굴인식 시스템에 대해 수행한 기만공격 실험을 설명하며, 마지막 4장에서는 결론을 맺는다.

Ⅱ. 적대적 공격

본 장은 적대적 공격을 공격 목표, 공격자의 지식에 따라 크게 두 가지로 분류하고 적대적 공격 방법을 설명한다.

2.1. 공격 목표에 따른 분류

적대적 공격으로 생성된 적대적 예제는 딥러닝 모델

* 공주대학교 대학원 융합과학과 (gsryu1026@smail.kongju.ac.kr)
** 공주대학교 산학협력단 전임연구원 (hspark@kongju.ac.kr)
*** 공주대학교 의료정보학과, 교신저자 (sunchoi@kongju.ac.kr)

이 어떤 class로 분류하느냐에 따라서 표적 공격 (targeted attack)과 무표적 공격(untargeted attack)으로 구분할 수 있다. 먼저, 표적 공격은 적대적 공격을 통해 생성된 적대적 예제를 딥러닝 모델이 원본 class가 아닌 공격자가 의도한 class로 인식하게 하는 공격을 말한다. [12] 예를 들어, 손글씨 이미지 데이터 세트인 MNIST (Modified National Institute of Standards and Technology)를 분류하는 문제에서 공격자는 적대적 공격을 이용하여 숫자 3에 해당하는 이미지를 딥러닝 모델이 숫자 3이 아닌 공격자가 의도한 숫자 9로 오인식하는 적대적 예제를 생성하여 공격하는 것을 말한다.

반면, 무표적 공격은 딥러닝 모델이 숫자 3이 아닌 임의의 다른 숫자로 오인식하는 적대적 예제를 생성하여 공격하는 것을 말한다. 일반적으로 무표적 공격은 표적 공격보다 적대적 예제를 생성하는 과정이 빠르고 변조되는 양도 적다. 하지만 표적 공격은 공격자가 원하는 class로 오인식을 유발할 수 있다는 점에서 무표적 공격보다 좀 더 정교하다. [13]

2.2. 공격자의 지식에 따른 분류

공격자가 공격 대상 딥러닝 모델에 대해 알고 있는 정보량에 따라 화이트 박스 공격 (white box attack)과 블랙 박스 공격 (black box attack)으로 나눌 수 있다. 화이트 박스 공격은 공격자가 공격 대상 딥러닝 모델의 구조, 파라미터, 출력값 등과 같은 모든 정보를 아는 상태에서 공격하는 것을 말한다. 화이트 박스 공격은 현재까지 100%의 성공률로 적대적 예제를 생성할 수 있는 것으로 알려져 있다. 하지만 공격자가 공격 대상 모델에 대한 모든 정보를 획득하기에 어려움이 있다.

블랙 박스 공격은 공격자가 공격 대상 모델에 대한 어떠한 정보도 없는 상태에서 공격하는 것을 말한다. 따라서 블랙 박스 공격은 공격자가 공격 대상 모델에 대해서 오직 입력값에 대한 출력값(분류 정보)만 알 수 있는 환경에서 이뤄진다. 블랙 박스 공격은 화이트 박스 공격보다 더 현실적이지만 공격 성공률은 현저히 낮다.

2.3. 적대적 공격 방법

적대적 공격으로 적대적 예제 x'을 생성하는 대표적인 4가지 방법이 있다. 첫 번째 방법으로 L-BFGS [10]가 있다. L-BFGS는 유클리드 거리 (Euclidean distance)를 사용하여 원본 이미지 x와 비슷한 적대적 예제 x'을 생성한다. L-BFGS는 다음과 같이 정의된다.

$$\begin{aligned} &\text{minimize } c \bullet \| x - x' \|_2^2 + \text{loss}_{F,t}(x') \\ &\text{such that } x' \in [0,1]^n \end{aligned} \tag{1}$$

여기서, $\text{loss}_{F,t}$는 딥러닝 모델의 손실 함수 (loss function)를 의미하고, t은 목표 class를 의미한다. L-BFGS는 공격 성공률을 높이면서 원본 이미지 x와 거리 차이를 최소화하는 적절한 c값을 여러 번 반복하여 찾는다.

두 번째 방법으로 Fast Gradient Sign Method (FGSM) [14] 가 있다.

$$x' = x - \epsilon \bullet \text{sign}(\nabla \text{loss}_{F,t}(x)) \tag{2}$$

여기서, ϵ은 탐지할 수 없을 만큼 충분히 작은 값을 선택한다. FGSM은 목표 딥러닝 모델의 손실 함수에 대한 기울기 (gradient)를 이용하여 원본 이미지를 변조시킬 방향과 강도를 결정한다.

세 번째 방법으로 Deepfool [15] 이 있다. Deepfool은 유클리드 거리를 최적화는 무표적 공격 기법이다. L-BFGS보다 더 가까운 적대적 예제를 생성할 수 있다. Deepfool은 뉴럴 네트워크 (neural network)가 전부 선형 (linear)라고 가정하여 구성되었으며 간단한 선형 문제를 푸는 것으로 적대적 예제를 생성한다. 하지만 뉴럴 네트워크는 실제로 선형이 아니므로 이 과정을 두 번 반복한다.

네 번째 방법은 기존 L-BFGS보다 성능이 향상된 Carlini 공격 [16] 이다. Carlini 공격은 현재까지 알려진 딥러닝 모델에 대해서 100%의 화이트 박스 공격 성공률을 보인다. Carlini 공격 중 하나인 L_2공격은 다음과 같이 정의된다.

$$\text{minimize } \| x' - x \|_2^2 + \text{c} \bullet f(x') \tag{3}$$

여기서, f는 [16]에서 정의한 7가지의 목적함수 (objective function) 중 가장 성능이 좋았던 목적함수이고, 다음과 같이 정의된다.

$$f(x') = \max(\max\{Z(x')_i : i \neq t\} - Z(x')_t, -k) \quad (4)$$

여기서, $Z(x')$은 적대적 예제를 목표 딥러닝 모델에 입력하였을 때, 적대적 예제에 해당하는 출력 logits를 의미한다. L_2공격은 목표 딥러닝 모델의 손실 함수를 사용하는 L-BFGS와는 다르게 목표 class에 해당하는 logits 값이 원래 class에 해당하는 logits 값과의 차이가 최소화하기 위한 목적함수를 정의하여 적대적 공격을 수행한다. 또한, 공격 성공률을 높이면서 원본 이미지 x와 거리 차이도 최소화하는 적절한 c를 찾으며, 신뢰도(confidence value)를 반영하여 공격 성공률을 조절할 수 있다. 신뢰도를 증가시킬수록, $f(x')$에 대한 가중치가 증가하여 공격 성공률이 높아지지만, 적대적 예제 x'과 원본 이미지 x의 차이가 증가한다.

Ⅲ. 딥러닝 기반 얼굴인식 모델 기만공격 실험

본 장에서는 딥러닝 기반 얼굴인식 시스템에 대해 기만공격을 수행하기 위해 목표 얼굴인식 모델 학습, 변조 영역설정, 기만공격에 대한 실험 결과를 설명한다. 본 실험에서는 화이트 박스 공격을 수행하였으며 적대적 공격 방법으로 Carlini 공격 중 하나인 L_2 공격으로 적대적 예제를 생성하였다.

3.1. 얼굴 이미지 수집 및 전처리

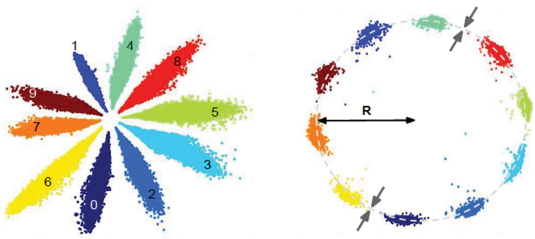
공격대상 얼굴인식 모델을 학습하기 위해, 연구실 9명 (남자 7명, 여자 2명)에 대해 하루에 50장 이상 12일 동안 얼굴 이미지를 수집하였다. 얼굴 이미지는 각도, 조명, 표정 등 다양한 변화를 주며 수집되었으며, 각자 사용하고 있는 스마트폰을 통해 사람마다 약 500장의 얼굴 이미지가 수집되었다. 수집된 얼굴 이미지에서 얼굴 추출 및 정렬하기 위해 사전에 학습된 Multi-task CNN (Convolutional Neural Network) [17] 을 사용하였으며, 전처리한 얼굴 이미지의 크기는 224×224로 통일시켰다. 전처리한 얼굴 데이터 세트에 대해 얼굴 특징점인 랜드마크 (landmarks) 정보를 추출하기 위해 FAN (Face Alignment Network) [18] 을 사용하였으며, 눈썹과 턱에 해당하는 랜드마크 점들을 연결하여 얼굴만 존재하는 또 하나의 얼굴 이미지 데이터 세트를 만들었다. 얼굴만 존재하는 이미지 데이터 세트를 만든 이유는 딥러닝 기반 얼굴인식 모델이 얼굴이 아닌 다른 부분을 보고 분류하는 경우를 배제하기 위해서이다. 그림 1은 두 데이터 세트에 대한 예를 보여준다.

(a) (b)

〔그림 1〕 두 데이터 세트에 대한 예. (a) 배경을 포함한 얼굴 이미지 데이터 세트. (b) 얼굴만 존재하는 얼굴 이미지 데이터 세트

3.2. 공격대상 얼굴인식 모델 학습

전처리한 두 개의 얼굴 이미지 데이터 세트를 이용하여 두 개의 공격대상 얼굴인식 모델을 학습하였다. 전처리한 두 개의 얼굴 이미지 데이터 세트에 대해서 학습 데이터와 테스트 데이터로 7:3으로 나눈 후 학습 데이터에 대해서만 좌우 반전시켜 데이터를 증폭시켰다. 두 개의 공격대상 얼굴인식 모델은 50개의 레이어 (layer)를 가진 ResNet 버전 2 [19] 를 이용하여 학습하였으며 손실 함수로는 소프트맥스 크로스 엔트로피 (softmax cross entropy)에 $0.1\times$Ring loss [6]을 더하여 사용하였다. 그림 2는 소프트 맥스 크로스 엔트로피와 소프트맥스 크로스 엔트로피에 Ring loss를 추가로 사용하여 학습된 딥러닝 모델에서 추출된 특징 분포의 차이를 보여준다. 약간의 배경이 포함된 데이터 세트로 학습된 첫 번째 얼굴인식 모델은 99.77%의 정확도를 보였으며 얼



(a) (b)

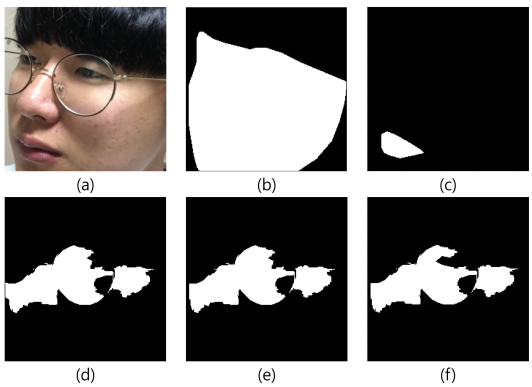
〔그림 2〕 MNIST 데이터 세트를 손실 함수로 (a) 소프트맥스 크로스 엔트로피와 (b) 소프트맥스 크로스 엔트로피 +Ring loss를 사용하여 학습된 딥러닝 모델에서 추출된 특징 분포에 대한 예 〔6〕

굴만 존재하는 데이터 세트로 학습된 두 번째 얼굴인식 모델은 99.8%의 정확도를 보였다.

3.3. 변조 영역 설정

변조 영역은 실제 얼굴에 분장 가능한 영역을 고려하여 설정하였다.

변조 영역을 제한하여 공격대상 얼굴인식 모델을 기만 공격하기 위해 두 가지 방법을 조합하여 변조 영역을 설정하였다. 첫 번째 방법은 FAN [18]을 사용하여 왼쪽 눈썹, 오른쪽 눈썹, 왼쪽 눈, 오른쪽 눈, 코, 입술, 턱에 대한 랜드마크를 탐지하여 활용하였다. 두 번째 방법은 설명 가능한 인공지능 (eXplainable Artificial Intelligence) 기술 중 하나인 LIME (Local Interpretable Model-agnostic Explanations) [20]를 활용하여 목표 얼굴인식 모델이 입력 이미지를 분류하는데 중요한 영역을 추출하였다. 두 가지 방법을 이용하여 총 5가지의 변조 영역을 설정하였다. 첫 번째로 얼굴 영역은 탐지된 랜드마크에서 눈썹과 턱에 해당하는 랜드마크 점들을 연결하여 설정하였고, 두 번째로 입술 영역은 입술에 해당하는 랜드마크 점들을 이용하여 설정하였다. 세 번째로 LIME 영역은 LIME에서 추출된 중요 영역을 말하고, 네 번째로 얼굴+LIME영역은 얼굴 영역과 겹치는 LIME 영역이며 마지막으로 얼굴+LIME 영역에서 눈과 눈썹에 해당하는 영역을 제외한 영역을 설정하였다. 눈과 눈썹에 해당하는 영역을 제외한 이유는 눈은 실제로 분장할 수 없고 눈썹 영역은 실제 분장하더라도 큰 차이를 보일 수 없어 제외하였다, 그림 3은 변조 영역설정에 대한 예를 보여준다.



〔그림 3〕 변조 영역설정에 대한 예. (a) 원본 이미지, (b) 얼굴 영역, (c) 입술 영역, (d) LIME 영역, (e) 얼굴+LIME 영역, (f) 눈과 눈썹 영역을 제외한 얼굴+LIME 영역

3.4. 실험 결과

테스트 데이터에서 사람마다 한 장의 얼굴 이미지를 선택하여 다른 8명으로 인식하도록 두 개의 얼굴인식 모델에 대해 표적 공격을 수행하였다. 표 1과 표 2는 두 얼굴인식 모델에 대해 변조 영역에 따른 기만공격 성공률을 비교한다.

얼굴 영역만 변조하여 생성된 적대적 예제로 기만공격하였을 때, 두 얼굴인식 모델 모두 100%의 공격 성공률을 보였다. 얼굴을 추출하여 정렬된 이미지이므로 이미지에서 얼굴은 상당히 넓은 영역을 차지한다. 따라서, 상당히 넓은 영역을 변조하여 적대적 예제를 생성하기 때문에 변조 여부를 파악하기 힘들고 공격 성공률 또한 높다.

입술 영역만 변조하여 생성된 적대적 예제로 기만공격하였을 때, 두 얼굴인식 모델 모두 공격 성공률이 현저히 떨어진 것을 볼 수 있다. 입술 영역은 얼굴 영역보다 좁은 영역을 변조하여 적대적 예제를 생성하기 때문에 공격 성공률이 낮고 변조 여부도 비교적 파악하기 쉽다.

〔표 1〕 배경을 포함한 얼굴 이미지로 학습된 얼굴 인식기에 대해 변조 영역에 따른 기만공격 성공률 비교

변조 영역	공격 횟수	공격 성공 횟수	공격 성공률
얼굴 영역	72	72	100%
입술 영역		2	2.78%
LIME 영역		72	100%
얼굴+LIME 영역		54	75%
얼굴+LIME 영역 (눈, 눈썹 제외)		53	73.61%

〔표 2〕 얼굴만 존재하는 이미지로 학습된 얼굴 인식기에 대해 변조 영역에 따른 기만공격 성공률 비교

변조 영역	공격 횟수	공격 성공 횟수	공격 성공률
얼굴 영역	72	72	100%
입술 영역		6	8.33%
얼굴+LIME 영역		72	100%
얼굴+LIME 영역 (눈, 눈썹 제외)		71	98.61%

〔그림 4〕 두 얼굴인식 모델에 대해 변조 영역에 따라 생성된 적대적 예제의 예

LIME 영역만 변조한 기만공격은 배경을 포함한 얼굴 데이터 세트로 학습된 얼굴인식 모델에 대해서만 수행하였다, LIME 영역은 딥러닝 모델이 입력 이미지를 분류할 때 중요하다고 판단한 영역이기 때문에 LIME 영역만 변조하여 생성된 적대적 예제로 기만공격하였을 때 100%의 공격 성공률을 보였다.

두 얼굴인식 모델에 대해 얼굴 영역과 겹치는 LIME 영역을 변조하여 기만공격을 수행하였을 때, 공격 성공률의 차이가 존재하였다. 배경을 포함한 얼굴 데이터 세트로 학습된 얼굴인식 모델에 대한 공격 성공률은 75%를 보였으며 얼굴만 존재하는 데이터 세트로 학습된 얼굴인식 모델에 대한 공격 성공률은 100%를 보였다. 배경을 포함한 얼굴 데이터 세트로 학습된 얼굴인식 모델에 대해 LIME 영역을 추출하였을 때, 몇 사람에 대한 얼굴 이미지를 분류할 때 주로 얼굴이 아닌 영역을 보고 분류하는 경향을 보였다. 따라서, 얼굴과 겹치는 LIME 영역이 좁거나 거의 존재하지 않아 공격 성공률이 떨어진 것을 볼 수 있었다. 얼굴만 존재하는 데이터 세트로 학습된 얼굴인식 모델은 모든 사람을 얼굴만 보고 분류하기 때문에 얼굴 영역과 겹치는 LIME 영역은 비교적 넓어 공격 성공률이 떨어지지 않았다.

눈과 눈썹 영역을 제외한 얼굴과 겹치는 LIME 영역을 변조하여 기만공격을 수행하였을 때, 두 모델에 대한 공격 성공률은 큰 차이를 보이지 않았다. 그림 4는 두 얼굴인식 모델에 대해 설정된 변조 영역에 따라 기만공격에 성공한 생성된 적대적 예제의 일부를 보여준다.

Ⅳ. 결 론

본 논문에서는 적대적 공격을 통해 생성된 적대적 예제를 이용하여 딥러닝 기반 얼굴인식 모델에 대해 기만공격을 수행하였으며 실제 얼굴에 분장할 수 있는 영역을 고려하여 설정된 변조 영역에 따른 기만공격의 성능을 분석하였다.

배경을 포함하는 얼굴 데이터 세트로 학습된 얼굴인식 모델은 일부 사람에 대해서 주로 얼굴이 아닌 다른 영역을 보고 분류하는 경향을 보여 얼굴 영역과 겹치는 LIME 영역 변조에 따른 기만공격 성공률이 떨어지는 모습을 보였다. 얼굴만 존재하는 얼굴 데이터 세트로 학습된 얼굴인식 모델은 입술 영역 변조를 제외한 나머지 영역 변조를 통한 기만공격은 높은 공격 성공률을 보였다. 또한, 적대적 공격을 통해 생성된 적대적 예제는 변조 영역이 좁을수록 변조 여부를 파악하기 쉬웠으며 공격 성공률도 현저히 떨어지는 모습을 보였다.

참 고 문 헌

[1] 김형일, 문진영, 박종열, “딥러닝 기반 고성능 얼굴인식 기술 동향”, 전자통신동향분석, 33(4), pp. 43-53, 2018

[2] M. Wang, W. Deng, “Deep Face Recognition: A Survey,” arXiv preprint arXiv:1804.06655, 2018.

[3] W. Liu, Y. Wen, Z. Yu, and M. Yang, “Large-Margin Softmax Loss for Convolutional Neural Networks,” Proceedings of the 33rd International Conference on Machine Learning,

pp. 507-516, 2016.

[4] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep Hypersphere Embedding for Face Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 212-220, 2017.

[5] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, and W. Liu, “CosFace: Large Margin Cosine Loss for Deep Face Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 5265-5274, 2018.

[6] Y. Zheng, D.K. Pal, and M. Savvides, “Ring loss: Convex Feature Normalization for Face Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 5089-5097, 2018.

[7] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” arXiv preprint arXiv:1801.07698, 2018.

[8] G. Hinton, L. Deng, D. Yu, G. Dahl, A.R. Mohamed, N. Jaitly, and T. Sainath, “Deep Neural Networks for Acoustic Modeling in Speech Recognition,” IEEE Signal Processing Magazine, 29(6), pp. 82-97, 2012.

[9] S. Potluri and C. Diedrich, “Accelerated Deep Neural Networks for Enhanced Intrusion Detection System,” 2016 IEEE 21st International Conference on Emerging technologies and Factory Automation, pp. 1-8, 2016.

[10] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, “Intriguing Properties of Neural Networks,” arXiv preprint arXiv:1312.6199, 2013.

[11] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, and D. Song, “Robust Physical-World Attacks on Deep Learning Visual Classification,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 1625-1634, 2018.

[12] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z.B. Celik, and A. Swami, “The Limitations of Deep Learning in Adversarial Settings,” 2016 IEEE European Symposium on Security and Privacy, IEEE, pp. 372-387, 2016.

[13] 권현, 윤현수, 최대선, “Evasion attack에 대한 인공지능 보안 이슈,” 정보과학회지, 36(2), pp. 32-36, 2018.

[14] I. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” arXiv preprint arXiv:1412.6572, 2014.

[15] S.M Moosavi-Dezfooli, A. Fawzi, and P. Frossard, “Deepfool: a Simple and Accurate Method to Fool Deep Neural Networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 2574-2582, 2016.

[16] N. Carlini, D. Wagner, “Towards Evaluating the Robustness of Neural Networks,” 2017 IEEE Symposium on Security and Privacy, IEEE, pp. 39-57, 2017.

[17] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks,” IEEE Signal Processing Letters, 23(10), pp. 1499-1503, 2016.

[18] A. Bulat and G. Tzimiropoulos, “How Far are We from Solving The 2D & 3D Face Alignment Problem?(and a Dataset of 230,000 3D Facial Landmarks),” Proceedings of the IEEE International Conference on Computer Vision, IEEE, pp. 1021-1030, 2017.

[19] K. He, X. Zhang, S.Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” European Conference on Computer Vision, Springer, pp. 630-645, 2016.

[20] M.T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You?: Explaining the Predictions of Any Classifier,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM,

pp. 1135-1144, 2016.

〈저 자 소 개〉

류 권 상 (Gwonsang Ryu)

학생회원

2016년 2월 : 공주대학교 응용수학과 학사

2018년 2월 : 공주대학교 대학원 융합과학과 석사

2018년 3월 ~ 현재 : 공주대학교 대학원 융합과학과 박사과정

<관심분야> 인증, 이상거래탐지, 머신러닝

박 호 성 (Hosung Park)

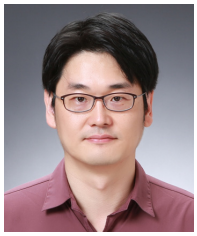
2008년 2월 : 충남대학교 컴퓨터공학과 학사

2014년 8월 : 충남대학교 컴퓨터공학과 석박사 통합

2014년 9월~2017년 4월 : 충남대학교 소프트웨어연구소 전임연구원/경상대학교 무인항공기SW 플랫폼연구사업팀 선임연구원

2017년 8월~현재 : 공주대학교 산학협력단 전임연구원

<관심분야> 인증, 정보보호, 머신러닝



최 대 선 (Daeseon Choi)

종신회원

1995년 2월 : 동국대학교 컴퓨터공학과 학사

1997년 2월 : 포항공과대학교 컴퓨터공학과 석사

2009년 1월 : 한국과학기술원 전산학과 박사

1997년 1월~1999년 6월 : 현대정보기술 선임

1999년 7월~2015년 8월 : 한국전자통신연구원 인증기술연구실 실장/책임연구원

2015년 9월~현재 : 공주대학교 의료정보학과 부교수

2016년 ~ 현재 : 정보보호학회 이사

<관심분야> 인증, 개인정보보호, 이상거래탐지, 의료정보보안, 머신러닝