자격시험에서 오토인코더 및 Semi-Supervised GAN 기반의 응시자 본인 확인 시스템 제안

임세진*, 김현지**, 강예준*, 김원웅*, 송경주**, 양유진**, 오유진**, 장경배**, 서화정**
*한성대학교 컴퓨터공학부
**한성대학교 IT융합공학부
dlatpwls834@gmail.com, etus1211@gmail.com, dnjsdndeee@gmail.com, khj1594012@gmail.com, thdrudwn98@gmail.com, yujin.yang34@gmail.com, oyj0922@gmail.com, starj1023@gmail.com, hwajeong84@gmail.com

Autoencoder and Semi-Supervised GAN-based candidate identity verification system in qualifying examination

Se-Jin Lim*, Hyun-Ji Kim**, Yea-Jun Kang*, Won-Woong Kim*, Gyeong-Ju Song**, Yu-Jin Yang**, Yu-Jin Oh**, Kyung-Bae Jang**, Hwa-Jeong Seo**
*Dept. of Computer Engineering, Hansung University
**Dept. of IT Convergence Engineering, Hansung University

요 약

국내에서는 매년 많은 수의 자격시험이 치러지고 있다. 현재 대부분의 시험장에서 응시자 본인 확인 절차는 감독관이 응시자의 얼굴과 신분증 사진을 비교하는 방식으로 이루어진다. 하지만 이 방식은 사람에 따라 오차가 클 수 있으며, 사진과 눈에 띄는 차이가 없으면 동일인물로 판단하기 쉽다. 최근까지도 대리응시 이슈가 발생하고 있어 근절을 위한 보다 강력한 조치가 필요하다. 본 논문에서는 지문과 오토인코더, SGAN을 이용하여 대리응시방지를 강화할 수 있는 본인 확인 시스템을 제안한다. 이때 응시자의 지문정보가 그대로 인증 서버에 저장되면 응시자의 생체정보가 노출될 수 있다는 문제점이 존재한다. 따라서 오토인코더를 통해 지문의 특징점만 추출하여 인증용 이미지를 생성하고 이 이미지를 서버에 저장하여 학습시키도록 한다. 적은 학습데이터 환경에서 분류기로써 좋은 성능을 갖는 SGAN을 통해 인증 이미지와 응시자가 동일인물인지 확인할 수 있다. 서버가 공격을 받더라도 응시자의 지문데이터가 그대로 노출되지 않게 되어 보안 취약점을 극복할 수 있다.

1. 서론

취업, 전문성, 학업, 자기계발 등의 목적으로 국내에서는 매년 많은 수의 자격시험이 치러지고 있으며, 다양한 연령층이 자격시험에 임하고 있다. 현재 대부분의 시험장에서의 응시자 본인 확인 방법은 감독관이 응시자의 얼굴과 신분증 사진을 비교하는 방식으로 이루어지고 있다. 하지만 이 방식은 개인의 판단에 맡기는 것이기 때문에 사람에 따라 오차가 클 수 있으며, 사진과 눈에 띄는 차이가 없으면 동일인물이라고 판단하기 쉽다. 필체를 비교하여 본인 확인을 강화하고 있지만, 최근까지도 시험 대리응시 이슈가 발생하고 있어 이 방식에도 한계점이 있다 [1,2]. 자격시험의 결과에 따라 승진, 취업, 진학 등에 영향을 미치는 민감한 사안인 만큼 대리시험의 근절을 위한 보다 강력한 조치가 필요한 상황이다.

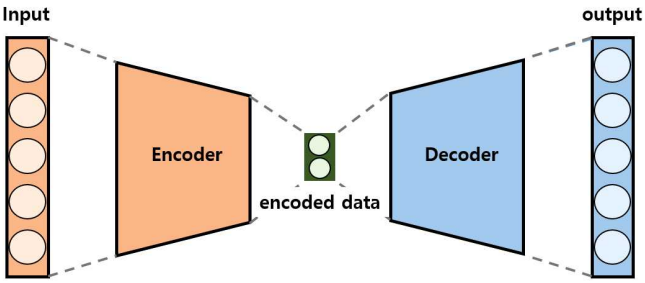
본 논문에서는 지문과 오토인코더, SGAN을 이용하여 대리응시방지를 강화할 수 있는 응시자 본인 확인 시스템을 제안한다. 이때 인증을 위한 지문데이터가 시험기관의 서버에 그대로 저장될 경우, 서버가 보안 위협을 받았을 때 응시자의 지문정보가 그대로 노출된다는 문제점이 있다. 이를 극복하기 위하여 오토인코더와 SGAN을 사용함으로써 지문데이터가 그대로 노출되지 않으면서도 동일인임을 확인할 수 있도록 하였다.

2. 관련 연구

2.1 오토인코더 (Auto Encoder)

오토인코더는 입력값과 출력값이 유사해지도록 데이터 레이블 없이 학습하는 비지도학습 형태의 인공신경망 모델이다. 오토인코더의 기본 구조는 아래 (그림 1)과 같다. 인코더(Encoder)는 입력된 데이터의 핵심 특징만을 추출하고, 은닉층에서 이 추출한 특징을 학습시키면 추출된 특징값을 바탕으로 디코더(Decoder)에서 원본 데이터와 근사한 값이 나오도록 재구성해준다. 즉, 인코더의

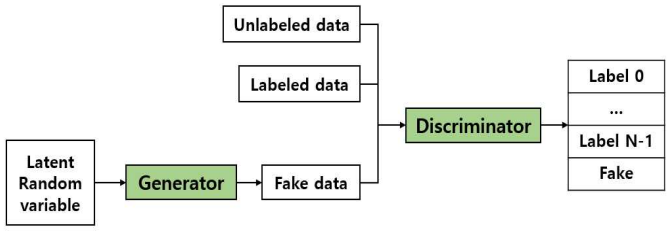
출력값은 입력데이터에서 불필요한 정보가 제거된 핵심 특징만이 추출된 형태임을 알 수 있다. 이때, 인코더와 디코더에 들어가는 노드수보다 은닉층에 들어가는 노드수가 더 적은 손실 압축 방법을 사용한다. 이러한 특성은 데이터 압축, 차원 축소(dimensionality reduction), 사전학습(pre-training), 잡음제거(denoising) 등에 활용된다[3, 4].



(그림 1) 오토인코더 구조

2.2 Semi-Supervised GAN (SGAN)

GAN은 생성자(Generator)와 판별자(Discriminator) 역할을 하는 2개의 network 모델이 서로 경쟁하면서 훈련 데이터와 구별이 안 될 정도로 훈련 데이터 셋의 특징이 잘 나타나는 샘플을 생성한다. 학습이 성공적으로 완료되었다면 좋은 성능의 생성자와 판별자를 얻게 된다. 주로 이미지 생성 및 복원 등 생성 모델로서 생성자가 많이 활용된다. GAN은 학습이 불안정하여 종종 의미없는 이미지를 출력하기도 하는데, 이런 점을 보완하기 위해 CNN과 GAN을 결합한 DCGAN(Deep Convolutional Generative Adversarial Network)이 제안되었다[5, 6]. DCGAN은 비교적 안정적으로 학습이 진행된다. Semi-Supervised GAN은 2016년에 ICML에서 제안된 기법으로, SGAN이라고도 하며, original GAN보다 성능이 좋은 DCGAN을 기본 구조로 갖는 모델이다[7]. 준지도 학습(Semi-Supervised learning)은 라벨이 있는 데이터를 이용한 지도 학습과 라벨이 없는 비지도 학습을 결합하여 상호보완을 통해 학습 능력을 개선하는 기법이다. 기존의 GAN을 활용한 모델들은 판별자가 단순히 진짜 데이터(Real Data)와 가짜 데이터(Fake Data)를 분류하는 모델로 사용되었지만, SGAN은 아래의 (그림 2)처럼 여러 클래스에 대해 분류를 수행하여 라벨링을하고 가짜 데이터를 구별할 수 있다. 즉, 판별자가 분류기 역할도 하는 것이다. 기존의 GAN이 좋은 성능의 생성자를 얻는 것이 목적이었다면, SGAN은 생성자와 판별자의 경쟁 학습을 통해 기존 GAN보다 좋은 성능의 판별자를 얻는 것에 초점이 맞춰져있다. 판별자의 분류 정확도는 학습 데이터가 100개 이하일 때 CNN(Convolutional Neural Network)보다 높았고, 1000개 이상으로 많아지면 CNN과 성능이 거의 같았다[7].

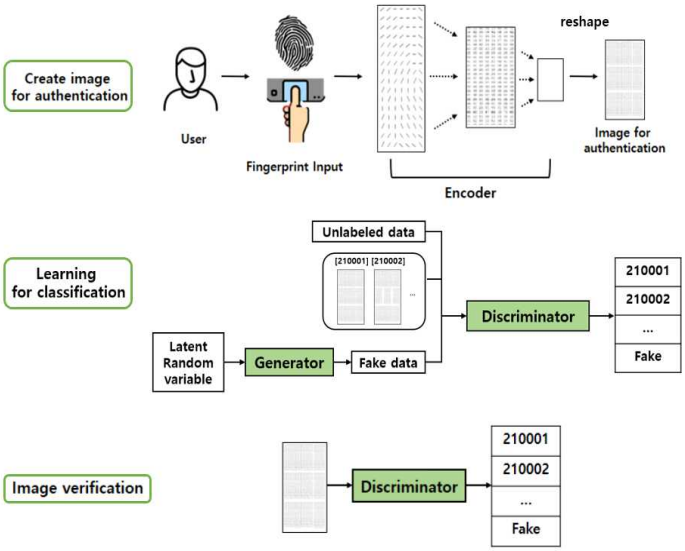


(그림 2) SGAN 구조

3. 시스템 제안

3.1 시스템 설계 구조

본 시스템은 (그림 3)처럼 오토인코더의 인코더를 이용하여 응시자의 지문데이터를 입력받아 특징점을 추출하여 재구성하는 인증용 이미지 생성 단계, 응시자로부터 인증용 이미지를 받아 분류기로 사용할 SGAN을 학습시키는 단계, 학습이 끝난 SGAN의 판별자를 이용하여 인증용 이미지가 입력되었을 때 응시자별로 분류할 수 있도록 하는 검증 부분으로 구성된다.



(그림 3) 시스템 구조

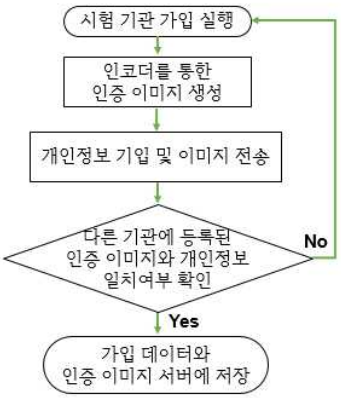
각 단계에 대해 더 자세히 설명한다. 먼저 인증용 이미지 생성 단계에서 주목할 점은 학습이 완료된 오토인코더에서 인코더만을 사용한다는 것이다. 오토인코더는 인코더를 통해 입력된 이미지의 주요 특징점을 추출하고, 해당 정보를 기반으로 하여 다시 원본 데이터로 복구하는 방식으로 동작한다. 하지만 본 시스템에서는 지문 데이터를 입력하여 원본과 비슷한 이미지를 생성하는 것이 목적이 아니라, 원본

데이터를 알아볼 수 없게 압축된 이미지 획득이 목적이므로 디코더 모델이 필요하지 않다. 따라서 지문 데이터에서 올바른 특징점을 추출할 수 있도록 학습을 하고, 이후 인증용 이미지 생성을 위해 사용할 때는 학습된 인코더만을 이용하여 특징점을 추출한다. 이러한 구조는 모바일 장치 상에 배포되는 모델의 용량을 줄일 수 있다[4].

다음 단계인 분류기 학습에서 분류기로 사용할 최종적인 모델은 SGAN의 판별자이다. SGAN은 주로 생성자의 좋은 성능에 초점을 맞춘 GAN 모델들과 다르게 판별자의 좋은 성능에 더 집중한 준지도 학습 모델이다. 또한 SGAN은 대표적인 이미지 분류모델인 CNN과 비교하였을 때 100개 이하의 적은 데이터셋 환경에서 높은 성능을 보인다. 대부분의 자격시험은 각 고사실 별 응시자의 수가 100명 이하이며, 고사실 별로 본인 확인을 진행하기 때문에 SGAN은 본 시스템에 적합하다. SGAN 학습 시에는 응시자의 인증 이미지마다 응시자의 수험번호로 라벨링이 되어 있는 데이터셋을 학습시킨다. 수험번호마다 클래스가 분류되기 때문에 생성자도 보다 정교하게 가짜 데이터를 생성하고 판별자가 구분하면서 학습을 진행한다. 마지막 단계에서는 앞선 단계에서 학습이 완료된 판별자만을 이용하여 인증 이미지를 분류하게 된다.

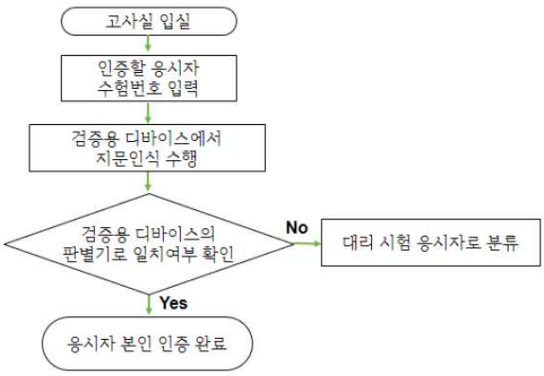
3.2 시스템 동작 시나리오

먼저 시험 응시자는 시험 접수 사이트에 가입 시에 배포된 인코더를 개인 모바일기기에 탑재하여 인증용 이미지를 생성하여 등록한다. 개인 모바일기기에서 인증용 이미지 생성을 한 후에 인증 서버로 전송하기 때문에 인증 서버는 응시자의 원본 이미지를 알 수 없다. 이때 지문인식을 지원하지 않는 모바일기기의 경우 자격시험 담당기관에서 등록하도록 한다. 시험기관의 서버에 가입자들의 인증 이미지가 보관되어있기 때문에 다른 응시자와 인증 이미지가 동일한 경우, 대리 응시시도를 의심할 수 있으며 이러한 방법을 통해 대리 응시시도를 사전에 차단할 수 있다. 또한 인증 이미지는 개인정보 노출에 대한 부담이 적기 때문에 시험기관별로 응시자의 인증 이미지를 공유하는 방법도 고려할 수 있다. 공유를 통해 동일한 인증 이미지로 가입되지 않은 경우, 즉, 응시자가 가입 당시에 대리인의 지문을 등록하는 것을 막을 수 있다.



(그림 4) 응시자 가입 흐름도

응시자가 시험을 접수하면 담당기관에서는 모든 응시자를 고사장, 고사실 별로 구분한 후, 인증데이터를 데이터셋으로 하여 SGAN을 각각 학습시켜 분류기 역할을 하는 판별자를 획득한다. 전체 응시자의 인증데이터를 고사장, 고사실 별 구분 없이 한꺼번에 학습시키게 되면 다른 고사장에서 시험을 치르는 대리인과 서로 바꿔서 응시할 수도 있으므로 방지하고자 하였다. 인증 과정에서 감독관이 수험번호를 입력한 후에 응시자의 지문인식을 통해 인증을 진행하도록 할 것이지만, 실제로 적용되었을 때는 응시생이 수험번호를 입력하고 인증을 하는 등의 다양한 변수가 존재할 수 있어 시스템적으로 예방하고자 하였다.



(그림 5) 자격시험장에서의 응시자 본인 인증 절차

자격시험장에서의 본인 확인 절차는 (그림 5)와 같다. 모든 응시자가 고사실에 입실을 마치고 고사실 밖으로 이동이 불가능한 상황에서 시험이 시작되기 전에 진행한다. 감독관은 각 고사실마다 배정된 디바이스를 사용하여 검증을 수행한다. 감독관은 각 응시자를 검증할 때마다 수험번호를 입력한 후, 검증용 디바이스에서 응시자의 지문인식을 수행한다. 검증용 디바이스 내의 인코더를 통해 생성된 인증

이미지는 판별자에 의해 수험번호에 해당하는 클래스로 분류되거나 Fake로 분류될 것이다. 수험번호와 같게 분류가 되었으면 응시자 본인 인증은 완료된다.

4. 결론

본 논문에서는 자격시험에서 대리응시를 방지할 수 있도록 지문인식, 오토인코더와 SGAN을 이용한 응시자 본인 인증 시스템을 제안하였다. 인코더를 통해 지문의 핵심 특징만 추출한 이미지를 사용함으로써 지문 전체 데이터를 노출하지 않으면서 본인 인증을 수행할 수 있도록 한 것에 의의가 있다. 신분증을 미소지한 경우 당일 시험에 응시할 수 없으며, 신분증을 분실한 성인이나 신분증이 없는 만 16세 미만의 응시자의 경우, 가족관계증명서, 신분확인증명서 등과 같은 추가적인 증명서류가 요구된다. 본 시스템을 통해 신분증 없이 모든 연령층에 같은 방식으로 본인 인증을 수행함으로써 앞서 말한 사항들에 유연하게 대처할 수 있다. 현재 검증 방식으로는 대리 시험을 진행했지만, 적발되지 않는 경우도 있을 수 있다. 하지만 본 시스템의 경우 응시자들의 인증 이미지를 서버에 보관하고 있기 때문에 다른 기관과의 인증 이미지 비교를 통해 이 전에 진행됐던 대리 시험을 적발할 수 있고 추후 발생할 수 있는 대리 응시를 방지할 수 있다. 본 논문은 자격 시험에서의 본인 인증을 위한 개념적 시스템을 제안하였으나, 시스템의 구체적인 구현이 향후 필요하다.

5. Acknowledgment

이 논문은 부분적으로 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2018-0-00264, IoT 융합형 블록체인 플랫폼 보안 원천 기술 연구, 50%) 그리고 부분적으로 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1F1A1048478, 50%).

참고문헌

[1] Moneys[Internet], https://moneys.mt.co.kr/news/mwView.php?no=2021042817438080756

[2] 한국일보[Internet], https://www.hankookilbo.com/News/Read/201810221667318032

[3] J.K.Lee, “A Noise-Tolerant Hierarchical Image Classification System based on Autoencoder Models”, Journal of Internet Computing and Services, pp.23-30, 2021.

[4] H.J.Kim, S.J.Lim, Y.J.Yang and H.J.Seo, "Privacy protection in user authentication using autoencoder and convolutional neural network", Korea Institute of Infomation Security&Cryptology, 2021.

[5] Radford.A, Metz.L and Chintala.S, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", ICLR 2016, arXiv:1511.06434, 2016.

[6] H.N.Jeon and D.H.Lee, "Wearable time series sensor data augmentation method through DCGAN with mode switching structure", The Korea Society of Mechanical Engineers, pp170-171, Apr, 2021.

[7] Odena A, "Semi-Supervised Learning with Generative Adversarial Networks", ICML 2016, arXiv:1606.01583, June, 2016.