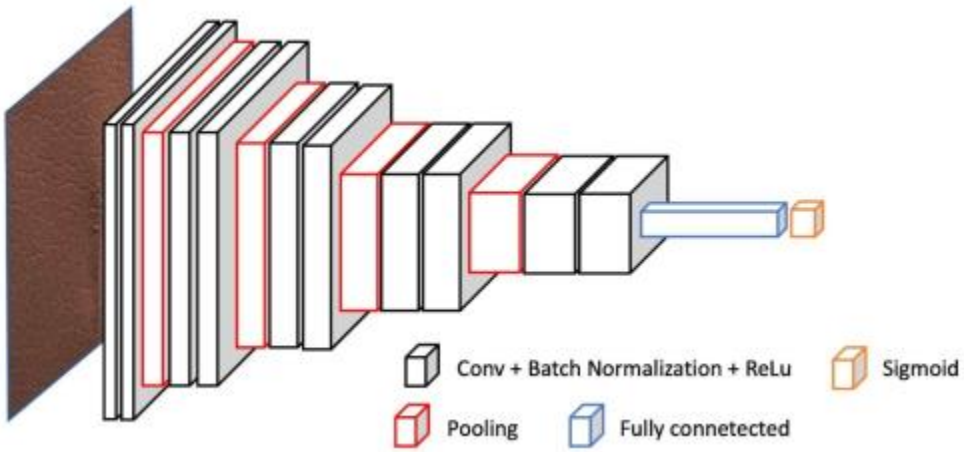
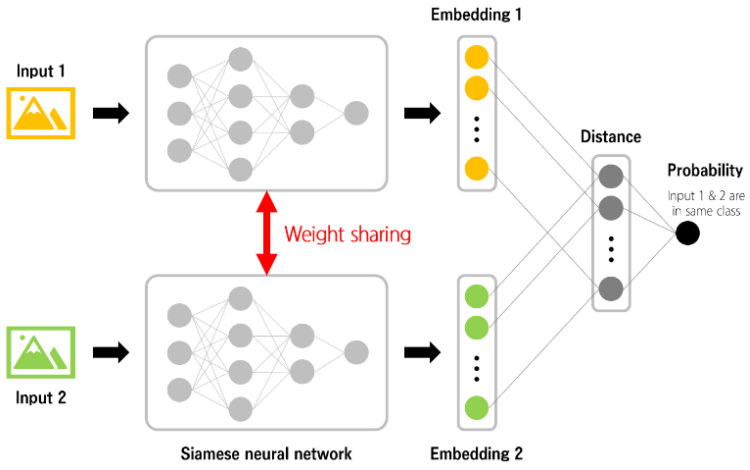
GAN 모델을 사용하는 이유

필자가 AI를 이용한 인증 시스템을 생각할 때, 생각한 모델이 세 가지가 있다.
첫 번째는 이진 분류 모델, 두 번째는 Siamese Network를 이용한 유사도 계산 모델, 세 번째는 GAN 기반 이상치 탐지 모델이다.



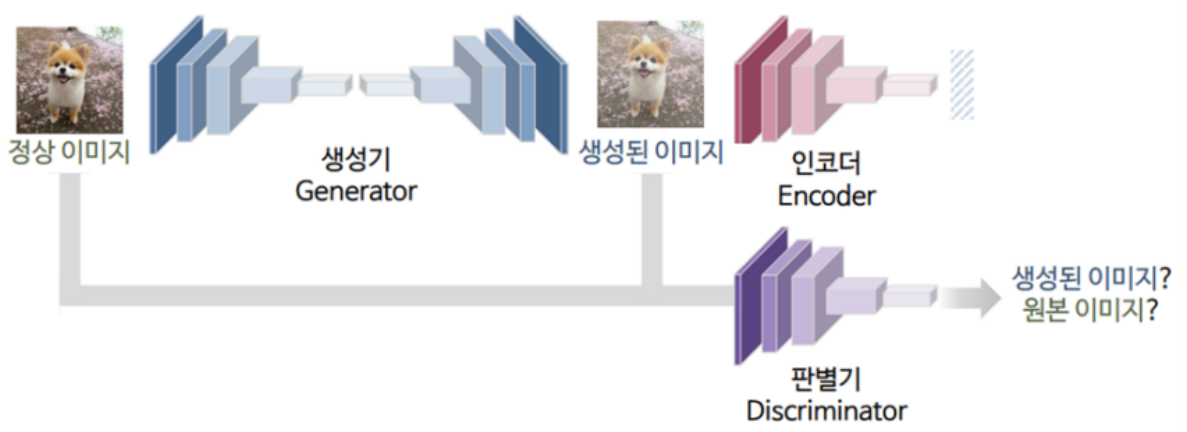
첫 번째 모델(이진 분류 모델)같은 경우, 정상 사용자의 사진이면 1을 출력하고 비정상 사용자의 사진(혹은 기타 사진)이면 0을 출력하는 학습을 한다. 이 모델은 매우 간단하다는 장점은 있지만 과연 다른 사용자와 제대로 비교를 할 수 있을지에 대한 의문점이 들었다. 왜냐하면 사람은 어느 정도 비슷한 생김새를 가지고 있기 때문에, 딥 러닝 모델로부터 나온 Feature Vector도 비슷할 것이기 때문이다. 각 사람의 Feature Vector에 대한 미세한 차이까지도 학습을 해야 정확한 결과가 나올 것이라고 생각했다.



두 번째 모델(Siamese Network를 이용한 유사도 계산 모델)같은 경우, 입력을 두 개 받아서 각각의 특징이 담긴 임베딩 벡터를 추출한 뒤, 두 벡터간의 차이(distance)를 이용해서 같은 클래스(1)인지 아닌지(0)를 예측하는 것이다. 이 모델은 Feature Vector간의 차이까지 고려하므로 첫 번째 모델의 문제점을 해결할 수 있겠지만 여전히 두 가지 문제점이 존재한다.

1. 이 모델은 유사한 입력과 유사하지 않은 입력 모두에 대해 학습할 필요성이 있다. 그런데 만약 데이터가 적으면 유사하지 않은 입력에 대해 학습시키기가 힘들다. 다른 사용자와의 차이를 학습시켜야 되는데, 다른 사용자 데이터가 없으면 학습시킬 수가 없다.

2. 반드시 두 개의 입력이 사용된다. 테스트를 할 때 첫 번째 입력은 서버에 저장된 정상 사용자의 사진이고 두 번째 입력은 사용자가 인증할 때 찍은 사진일 것이다. 따라서 서버에는 항상 사용자의 사진이 저장되어 있어야만 한다. 이는 두 가지 문제점을 초래한다. 첫 번째는 대용량의 저장 장치에 대한 비용 문제이다. 두 번째는 서버 해킹 시, 사용자의 정보가 유출될 수 있다는 문제이다. 따라서 보안 인증 시스템에는 적합하지 않다고 생각했다.



세 번째 모델(GAN 기반 이상치 탐지 모델)같은 경우, 정상 사용자 사진만을 잘 생성하는 모델을 학습시켜서, 입력 이미지와 생성된 이미지의 차이가 threshold 이상일 때 비정상이고, 그렇지 않을 때 정상이라고 출력하는 방식이다.

정상 사진만으로 학습되었기 때문에, 만약 비정상 사진이 입력으로 들어온다고 해도 생성된 이미지는 정상 사진과 유사할 것이다. 따라서 비정상 입력과 생성 입력 간의 차이가 발생한다.

이 모델이 두 번째 모델보다 성능이 높을지는 실험을 해봐야 알 수 있다. 다만 이 모델은 분류 문제를 대신해서 풀 수 있을 정도로 준수한 성능을 보이고, 무엇보다 두 번째 모델의 문제점을 완벽하게 해결한다.

1. 이 모델은 정상 입력만으로 학습된다. 따라서 다른 사용자 데이터가 없더라도 학습시키는 것에 문제점이 없다.

2. 이 모델은 한 개의 입력만이 사용된다. 실제 테스트 시, (사용자 인증할 때 찍은 사진)만을 입력으로 받는다. 따라서 각 사용자의 정보를 서버에 저장할 필요가 없다.

결국 학습의 효율성과 보안을 위해서는 'GAN 기반 이상치 탐지 모델'이 가장 적합하다고 생각하였다. 다만 생성모델이 꼭 GAN일 필요는 없다. 필자는 GAN을 이용한 방식이 많이 발전되었기 때문에 이렇게 제안하는 것이다.