특집 04

빅데이터 환경에서 얼굴의 스푸핑 공격 방지 기법을 위한 알고리즘의 설계

김동준 (한국과학기술원)

목 차

1. 서 론

2022년 현재 코로나 시대와 함께, 전 세계는 비대면 시대를 걸어가고 있으며 그에 대한 신속하고 정확한 얼굴 인식(face recognition) 기술이 많이 활용되고 있다. 사람의 얼굴을 보고 체온을 측정할 뿐만 아니라, 현재 어떤 시설을 사용할 때 혹은 건물을 입장할 때 등 얼굴 인식 기술은 갈수록 중요성이 커지고 있으며 이를 악용할 처지에 대한 여부도 증가할 가능성이 있다. 또한, 빅데이터 환경이 갖추어져 우리의 생활에는 많은 얼굴 데이터가 빅데이터로 형성되어 있다. 얼굴의 체온측정, 얼굴의 각 객체 인식, 그리고 얼굴 인식을 통한 보안시스템 등으로 많은 활용이 가능하며 이에 대한 악용의 우려도 증가하였다. 따라서, 얼굴 인식의 보안과 안정성을 위하여 얼굴 인식의 발달요구도 마찬가지로 증가했다. 이런 요구의 증가와 함께, 얼굴 인식 공격 방지 기법을 위한 알고리즘을 설계하고자 했다. 이 기술은 2021년도 서울특별시 산학연 협력사업 중 '제4회 서울 혁신챌린지(결선)' 인공지능 학습 기반의 얼굴 스푸핑 공격 방지 기법 알고리즘 기술개발의 과제(IC210001, 기여율 1/1)에 지원받아 개발된 기술이다. 한편, 본원 발명의 모든 측면에서 서울특별시의 재산 이익은 없다.

2. 얼굴 인식의 발달과 발전

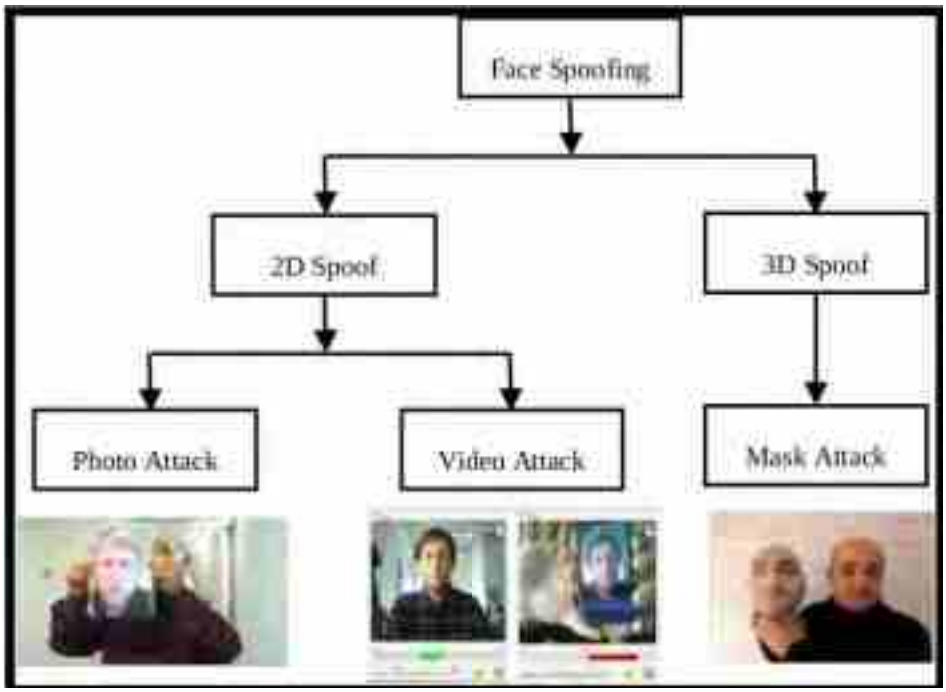
얼굴 인식 기술은 다양한 형태로 발전해 왔다. 크게는 다수의 사람이 있는 이미지에서 각 사람을 분류하는 형태와 한 사람의 이미지를 인식 및 분류(classification)하는 형태로 나눌 수 있다. 처음의 얼굴 인식 기술의 개발 형태는 사람 자체를 찾는 느낌의 형태가 컸었으며, 이를 정확하게 사람이라고 판단할 수 있고 그 위치를 찾는 것에 집중되어 있었다. 단, 점점 보안성이 대두되면서 각 얼

굴이 주어질 때 이 사람이 누구인지 정확하게 인식 및 분류하는 형태로 나아갔다. 딥러닝(deep learning)의 발달로 이미지 하나를 통해 질감, 색감, 얼굴의 형태 등 자세한 정보를 취득하여 이를 기준으로 사람의 얼굴을 더 정확하게 분류해나가는 것이 가능해졌다. 초기에는 Support Vector Machine(SVM)과 같은 머신러닝 기법을 사용했지만, 다음에 딥러닝이 발전하면서 기본적인 Convolution Neural Network(CNN)[1]을 통하여 이미지를 처리 및 분류하게 되어 나갔다.

이렇게 단순히 얼굴 이미지를 인식 및 분류하는 것이 발달하면서 이를 속이려는 시도 또한 늘어나게 되었다. 자신의 얼굴이 아닌 촬영된 다른 사람의 얼굴을 활용하는 공격이 스푸핑 공격 기법의 종류 중 하나이며 가장 대표적인 공격 기법이다.

3. 스푸핑 공격 기법

앞서 말한 기법과 같이, 스푸핑 공격 기법에는 다양한 종류가 있으며 다음 (그림 1)과 같이 나눌 수 있다. 두가지로 2D와 3D 공격 기법으로 처음에 분류가 되며, 3D 공격 기법은 자체 제작된 3D 마스크를 활용하여 실제 인물인 척 속이는 얼굴 인식 공격 기법이다. 이러한 경우는 3D 마스크를 특수하게 제작해야 하므로 공격하기 힘든 형태로 나타날 수 있다. 2D 공격 기법으로는 크게 사진을 활용한 방법과 비디오를 활용한 방법이 있다. Photo Attack은 프린터기로 인화된 사람 얼굴의 사진을 사용하여 얼굴 인식을 속이는 공격 기법이고, Video Attack은 앞서 말한 것과 같이 촬영한 인물의 동영상을 사용하여 얼굴 인식을 속이는 공격 기법이다.



(그림 1) 스푸핑 기법의 종류

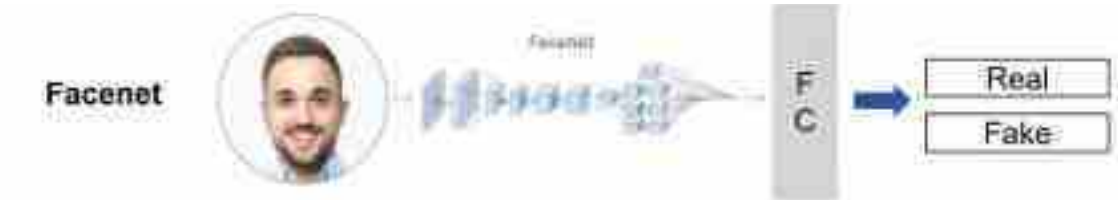
(그림 2) ResNet과 Densenet을 활용한 스푸핑 공격탐지 모델

가장 활용도가 높고 실제로 얼굴 인식 기술을 공격할 때 범용적인 부분은 Video Attack이며, 카메라 기술의 발달로 실제로 카메라에 비치는 얼굴의 모습과 촬영된 영상을 틀어둔 모습이 차이가 미미해지고 있다. 스푸핑 공격을 탐지하는 기술도 이와 맞추어 발전하고 있으며 CNN을 활용하여 얼굴의 진짜(Real)와 가짜(Fake)를 학습하여 이를 분류하는 일을 위주로 발전해 나가고 있다. CNN의 발전을 통하여 가장 간단한 네트워크인 SVM부터 현재 다양한 형태로 사용되고 있는 ResNet[2], Densenet[3] 등을 활용하여 진위 판단을 하는 성능은 점차 올라가고 있다.

4. 스푸핑 공격 탐지 알고리즘

기존에 존재하는 알고리즘들은 간단한 네트워크를 사용하여 분류를 해왔었고, 더 나아가선 ResNet, Densenet을 활용하는 방법이 있었다. 이는 확실히 효과가 있지만, 기존 ResNet과 Densenet은 원래 ImageNet이라는 1000가지의 이미지들을 분류하는 데이터셋을 위하여 학습된 네트워크이다. 물론 범용성을 봤을 때는 큰 문제가 없지만, 우리가 활용을 할 때에는 이 데이터 세트에 미리 학습된 ResNet과 Densenet을 가져와 얼굴 데이터 세트를 추가로 학습한 뒤 진짜와 가짜를 분류하게 된다. 이는 오히려 우리가 얼굴의 진위를 분류하는 특수성에는 맞지 않는다고 판단을 했다. 이를 해결하기 위하여, 다른 학습기법인 전이학습(Transfer learning)[4]을 활용했다. 전이학습이란, 기존의 다른 학습에 사용된 네트워크의 가중치(weight)를 가지고 와 이를 다른 학습의 기본 뼈대로 활용하는 학습기법이다. 이를 통하여 기존 가중치가 다른 학습에 사용되기 위하여 더 좋은 쪽으로 활용할 수 있게 하게 도와준다. 이를 활용하여 단순히 얼굴 인식만을 위해 학습되어온 네트워크인 FaceNet[5]을 가지고 와 사용하기로 했다. 따라서, 기존에 1000가지의 이미지로 학습되어온 가중치인 Imagenet을 사용하지 않고 조금 더 얼굴에 초점을 맞는 가중치인 FaceNet을 활용하여 기존뼈대를 구성하고, 분류하기 위한 Fully Connected(FC) layer를 부착하여 얼굴의 진위를 분류하기 위한 구성을 하였다. 이를 구성하였을 때의 모습은 (그림 3)과 같다.

진위 판단을 위한 학습을 위하여 자체적인 데이터를 구성하였다. 각 데이터는 동영상이 촬영된 진짜의 얼굴과(Real), 그 진짜의 얼굴을 전면 카메라로 촬영한 가짜 얼굴(Fake1), 후면 카메라로 촬영한 가짜 얼굴(Fake2), 그리고 칼러 프린트로 출력하여 촬영한 2D 얼굴 사진 (Fake3)로 구성하였으며 모두 1만 8천여 장의 이미지 데이터 세트를 구성하였다. 데이터 세트는 각각 70%, 15%, 15%를 학습, 검증, 시험용으로 사용하였다.



(그림 3) Facenet을 활용한 스푸핑 모델의 구조

또한, 데이터 세트를 충분한 양으로 구성했지만, 네트워크 학습의 안정성과 데이터가 증가한 효과를 내기 위하여 데이터 증강 기법[6]을 활용하였다. 이는 하나의 이미지를 회전, 기울이기, 수평 혹은 수직 뒤집기 등의 간단한 조작으로 하나의 데이터를 여러 가지 데이터처럼 활용하기 위한 기술이라고 볼 수 있다. 데이터 증강을 위하여, 이미지 확대 범위 15%, 기울이기 각도 0.2°, 회전 각도 20°, 수평 뒤집기만 존재 한 채로 진행하였다. 데이터 증강 기법을 활용하여, 동일 인물의 비슷한 사진을 여러 장면을 보게 되어 초래할 수 있는 오버피팅 문제를 어느 정도 없앨 수 있는 모습을 보여주었다.

최종적으로 전이학습을 위하여 기본 뼈대는 Facenet을 가지고 왔으며, 이를 통해 나온 이미지 feature를 가지고 두 개층의 FC layer로 구성된 분류기(classifier)로 가짜와 진짜를 판단하는 네트워크를 구성하였다. 이때, FC layer의 구성은 두 개의 층 구조를 이루게 하였으며, 각 512차원의 벡터를 256차원으로 줄고 최종적으로는 Real과 Fake 두 개로 분류를 하기 위해 2차원의 벡터로 축소하는 형태로 구성하였다. 이때의 손실 함수는 Binary cross entropy를 사용하였고, 최적화 함수로는 확률적 경사 하강법을 사용하였다. 이때의 learning rate는 0.001이었다.

얼굴 스푸핑 공격 검술의 성능 확인을 위하여 정확도와 precision, recall, f1 score를 비교하였다. 이를 통하여 최종적인 정확도는 95.32%의 성능을 냈다. 이때, 각각의 precision, recall, F1 score는 0.945, 0.935, 0.935로 나왔으며 각 이미지를 분류하는 알고리즘의 속도는 0.1초 이내로 해결하는 성능을 보여주었다. 각 성능을 기존의 모델인 ResNet과 Densenet을 활용하여 얼굴의 진위 판단했을 때의 성능 비교는 아래 <표 1>과 같다. 이때 precision과 recall 모두 한 쪽이 기울어진 형태로 나타나지 않았으며 이를 통하여 F1 score도 안정적으로 나온 결과를 보여주었다.

(그림 4) 데이터 세트의 종류

〈표 1〉 ResNet, Densenet, Facenet을 사용했을 때의 각 성능 비교

	ResNet	Densenet	Facenet
정확도	88.75%	90.78%	95.32%
Precision	0.91	0.92	0.945
Recall	0.895	0.9	0.935
F1 Score	0.905	0915	0.935

5. 결 론

앞에서 서술한 바와 같이, 얼굴 인식의 중요성은 증가하였으며 이에 대한 보안 문제의 중요성도 증가하였다. 이를 해결하기 위하여 딥러닝의 발전을 통한 스푸핑 공격 기법 감지 기술이 개발되었다. 기존의 스푸핑 공격 기법은 딥러닝에서 좋은 성능을 보이는 모델을 가지고 와 단순히 분류하는 데 활용하였지만, 얼굴의 분류라는 특수성에 초점을 맞추기 위하여 얼굴 분류를 위한 네트워크를 가지고 와 전이학습에 활용하였다. 이를 통하여, 얼굴 인식에 미리 학습된 네트워크를 통해서 얼굴의 분류를 위한 새로운 학습을 더 잘 받아들인다는 것을 확인하였다. 또한, 적은 데이터로 더 많은 데이터처럼 활용하기 위하여 데이터 증강 기법을 사용하였다. 동영상으로부터 추출된 사진데이터는 데이터의 중복성을 초래하여 오버피팅의 문제를 일으켜 모델의 안정성을 저하할 수 있지만, 데이터 증강 기법을 통하여 이를 해소할 수 있는 점을 확인했다. 이를 통해 기존 ResNet과 Densenet만을 사용한 모델들 보다, 더욱 안정적이고 성능이 향상된 학습 결과를 얻을 수 있었다.

더욱 추가하여 발전할 요소로는, 3D depth map을 활용하여 정말 가짜와 진짜에 대한 질감 색감뿐만 아니라 얼굴의 실질적 입체 정보도 주어지는 경우다. 화면으로 비치는 얼굴이나 사진으로 인화된 얼굴은 2D 정보에서 그치지만, 우리 현실에서는 3D의 정보들이 주어지고 있고 이를 사용해야 하는 상황이 더욱 많다. 이러한 상황에 맞추어 3D depth map을 추가로 사용한다면, 이미지의 단순한 표면적인 정보뿐만 아니라 2D 얼굴과 3D 정보를 같이 활용할 수 있게 된다. 이렇게 입력단에서 정보를 추가할 수도 있지만, 네트워크를 더 고도화하고 다른 기술과 접목한다면 3D 얼굴에 대해서 현실적인 접근과 그에 관한 결과를 보여줄 수 있을 것이다.

참 고 문 헌

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS; Proceeding of the 25th International Conference on Neural Information Processing Systems - Volume 1, December 2012

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, December 2016

[3] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. Densely Connected Convolutional Networks, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017

[4] Weiss, K., Khoshgoftaar, T.M. & Wang, D. A survey of transfer learning, Journal of Big Data 3, Article number: 9, 2016

[5] Florian Schroff, Dmitry Kalenichenko, James Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. IEEE Conference on Computer Vision and Pattern Recognition,

pp.815-823, 2015
[6] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019)

저 자 약 력

김 동 준

이메일 : rassilon712@kaist.ac.kr

- 2015년 고려대학교 컴퓨터학과 (학사)
- 2020년~현재 한국과학기술원 인공지능대학원 (석박사 통합)
- 관심분야 : 컴퓨터 그래픽스, 인공지능