영상 기반 대면편취 보이스피싱 피해예방 방법 연구

김영민†

요 약

대면 편취형 보이스피싱 피해 사례가 증가하고 있다. 이는 기존엔 사람에 의해서만 예방이 가능하였다. 본 연구에서는 이러한 문제점을 해결하고자 ATM의 영상 기반을 통해서 출금자의 보이스피싱 피해 여부를 예측한다. 객체 탐지를 통해 출금자의 통화 여부 및 마스크 착용 여부를 판단하고, 표정 인식을 통해 출금자의 불안 여부를 예측한다. 또한, 출금자의 개인 금융 데이터를 이용하여 금융 거래 이상 탐지를 이용하여 더욱 정교한 보이스피싱 피해 위험도를 산출한다.

1. 서 론

2020년 보이스피싱(Voice Pshing)에 따른 피해액은 7000억원까지 치솟았다. 이는 과거에 비해 최고 피해액이다. 또한, 피해자 수도 증가하여 2020년 31,681명의 피해자가 발생하였다.

정부는 보이스 피싱 피해를 예방하기 위해 여러 대책을 강구하고 있다. 하지만 스미싱(Smishing) 위주의 예방책은 많지만, 대면편취로 인한 예방책은 부족한 상황이다. 대면편취는 피해자가 현금 인출 후 보이스 피싱 범죄자에게 대면으로 직접 인출한 금액을 전달하는 형태의 금융사기를 뜻한다. 이러한 유형의 피해는 보이스 피싱 발생 건수 중 대면편취 피해 비율이 2019년 8.6%에서 2021년 8월엔 73.8%로 급증하였다.

따라서 본 연구는 대면편취 피해사례가 주로 발생하는 ATM 기계를 중심으로 ATM 영상 기반 객체 인식 및 표정 인식에 '마이 데이터(My data)'를 이용하여 보이스피싱 위험도를 계산하고자 한다.

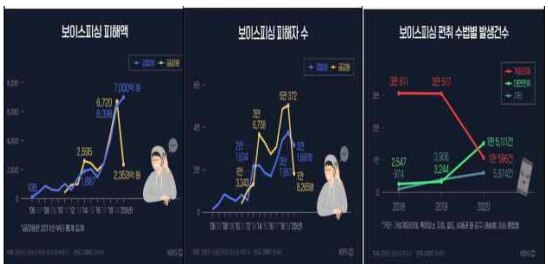

Fig. 1 보이스 피싱 현황

2. 관련 연구

2.1 YOLOv5

YOLOv5[1]는 YOLO(You Only Look Once)[2]를 기반으로 한 객체 탐지(Object Detection) 기술이다. 특히 FPS(Frames per Seconds)가 높아 빠른 처리 속도를 가진 장점이 있다. 그 중 가장 최근에 제안된 YOLOv5 알고리즘의 아키텍쳐(Architecture)는 C3로 (Fig. 2)와 같이 구성된다. 입력 이미지에서 특징을 추출하는 BackBone은 BottleNeck[3] 층과 CSPNet[4]을 결합하고 활성화 함수로 SiLU[5]를 사용하여 C3 구조를 이룬다.

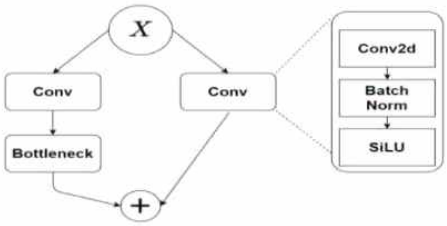

Fig. 2 C3 Architecture

COCO 데이터셋[6]을 이용하여 AP(Average Precision)과 FPS를 계산하였을 때 (Fig. 3)과 같은 결과가 나온다. 파라미터의 수가 가장 많은 x 모델은 mAP^{val} 68.9, GPU V100 기준 12.1ms 이다.

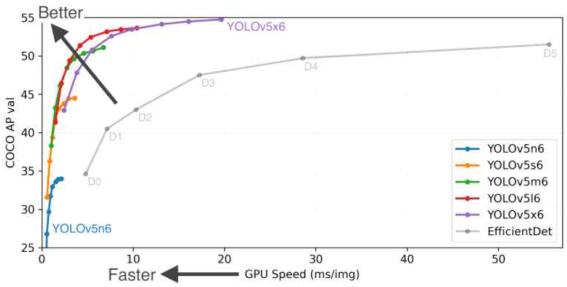

Fig. 3 Performance of YOLOv5

2.2 EfficientNet

EfficientNet[7]은 구글에서 만든 분류모델로 적은 파라미터(parameter)로 좋은 성능을 내는 모델이다. EfficientNet은 Compound Scaling이라는 개념을 적용하여 이미지의 스케일(scale)을 수식화하여 정확도와 처리 속도를 높였다.

EfficientNet은 파라미터의 수에 따라 b0부터 b7의 모델을 제안하고 정확도 및 FLOPs는 (Fig. 4)와 같다.

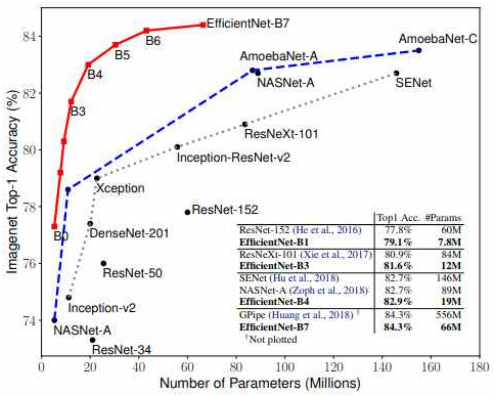

Fig. 4 Performance of EfficientNet

3. 제안 방법

본 연구에선 대면편취 피해를 예방하기 위하여 객체 탐지를 통해 전화 여부를 판단하고 표정 인식을 통해 출금자의 보이스 피싱 여부를 예측한다. 또한, COVID-19 상황에 맞춰 마스크를 탐지하고 벗어달라는 요청을 통해 더욱 정교한 표정 인식을 할 수 있게 한다.

표정 인식 후에 개인의 금융 거래 데이터를 이용하여 이상 출금 및 최근의 대출 여부를 통하여 이상 금융 거래인지 판단한다.

이후 출금자의 보이스피싱 피해 위험도를 계산한다. 이러한 과정은 (Fig. 5)와 같다.

2.1 객체 탐지

객체 탐지는 YOLOv5 모델을 이용하여 실시간 탐지를 이뤄낸다. 출금자의 마스크와 얼굴, 손을 탐지한다. 각각 훈련 데이터셋은 (Table 1)과 같다.

Table 1. Dataset

Object	Dataset Name	Number of data
Mask	Kaggle Mask Dataset[8]	mask : 500 no mask : 500
Hand	COCO-Hand[9]	hand : 25,000
Face	한국인 감정 데이터셋[10]	face: 50,000

여러 가지 데이터셋을 혼합할 경우 동일 클래스에 대한 annotation이 되어있지 않은 경우가 발생하여 직접 annotation을 시행하였다. 그리고 이 데이터를 COCO 데이터셋으로 미리 사전 학습된 yolov5m 가중치에 전이학습(transfer learning)을 시행하였다. 각 클래스별 mAP는 Mask는 0.9, Face는 0.8, Hand는 0.5로 나타났다.

전화 여부를 판단하기 위해 손의 bounding box를 이용한다. 얼굴 또는 마스크의 bounding box와 손의 bounding box가 겹칠 때 전화 행동으로 판단한다. 그리고 이러한 행동의 프레임이 30프레임 이상 지속되면 “전화 중”이라고 판단한다. 그리고 “전화 중”인 상태가 되면 위험도 가중치 w_{call}은 0.3으로 정의한다. “전화 중” 상태가 아니어도 위험도 가중치는 0.1로 정의하여 보이스피싱의 위험 가능성을 남겨둔다. 이는 식 (1)과 같이 나타난다.

$$call_n = \begin{cases} 1\ (A_{Face} \cap B_{Hand} \neq \Phi) \\ 0\ otherwise \end{cases} \quad (1)$$

$$w_{call} = \begin{cases} 0.3 \sum_{i \in frame}^{frame} call_i > 30 \\ 0.1\ otherwise \end{cases} \quad (2)$$

2.2 표정 인식

출금자의 불안한 상황을 탐지하기 위해 출금자의 표정 인식을 통해서 불안한 상황을 예측한다. 이를 위해 한국인에 특화된 표정 데이터셋을 이용하여 불안 또는 당황의 표정을 하나의 클래스로 위험 표정이라고 정의하였고, 이에 상반되는 중립 표정을 다른 클래스로 정의한다. 따라서 위험 또는 중립의 클래스로 이진 분류 문제가 된다.

실시간 표정 인식을 위하여 동영상 전체의 이미지가 아닌 주변 환경을 고려하지 않고, 사람의 표정만 잘라내어(cropping) 잘라낸 이미지만을 통해 표정 인식을 시행한다.

잘라낸 얼굴 이미지의 크기는 456x456 크기로 고정시키고, Optimzier는 AdamW[8] 방식을 사용하고 학습률 스케쥴러(Learning rate Scheduler)는 CosineAnnealing[9] 방식을 사용하였다. 중립 표정을 가진 500장의 사진과 위험 표정을 가진 500장의 사진을 학습 이미지로 사용하였고, 검증용 이미지는 각각 100장을 사용하여 총 200장의 이미지로 검증을 하였다. 표정 인식에 활용한 모델은 EfficientNet-b4를 사용하였다. 검증용 데이터셋의 평균 정확도는 약 91%이고, 에폭(epoch)별 정확도는 (Fig. 6)과 같다.

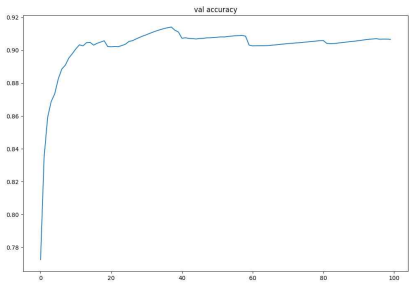

Fig. 5 Validation Accuracy(Facial Recognition)

표정 인식을 한 결과를 이진값(binary value, 0/1)으로 사용하지 않고, softmax를 적용한 결과를 이용하여 출금자의 표정이 “위험 표정”일 확률로 위험 가중치를 계산한다. 이는 식(3)과 같다.

$$w_{facial} = 0.3 \times P_{danger} \quad (3)$$

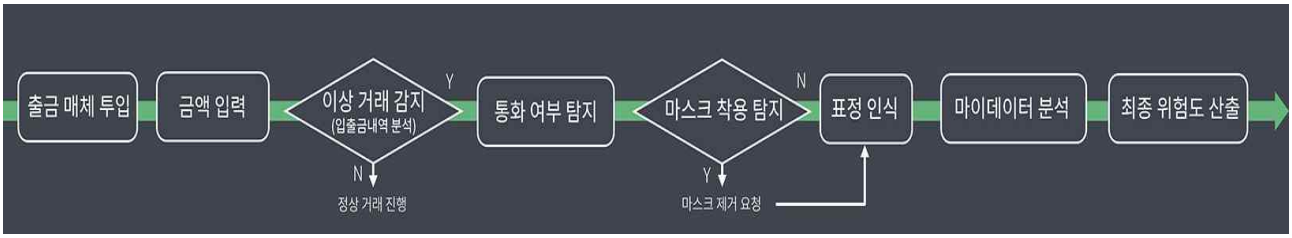

Fig. 5 Flow Chart

2.3 위험도 계산

보이스피싱 위험도는 금융데이터의 이상 여부도 고려한다. 개인 금융데이터의 이상 가중치는 식 (4)와 같다. x_1은 현재 출금금액의 이상 여부, x_2는 대출 정보, x_3는 카드 이용 정보의 이상 여부를 의미한다.

$$w_{bank} = 0.4 \times (x_1 + x_2 + x_3) \quad (4)$$

이상 여부는 통계적 이상치 정의에 따른다. 식 (5)의 $X_{withdraw}$는 현재의 출금 금액을 의미한다. IQR은 데이터의 Q_3(상위 75% 값) - Q_1(상위 25% 값)을 의미한다.

$$x_{withdraw} = \begin{cases} 0.4 & X_{withdraw} \geq Q_3 + IQR \times 1.5 \\ 0 & otherwise \end{cases} \quad (5)$$

대출 정보와 카드 이용 정보는 개인의 대출 및 지출 내역이 급격하게 증가한 경우를 이상 여부라 판단하고 모두 식 (6)과 같은 이상치 식을 이용한다. 그러나 x2, x3는 이상일 때 가중치를 0.3의 값을 준다. 이를 정리한 식은 식(6)과 같다.

$$x_1 = \begin{cases} 0.4 & abnormal \\ 0 & otherwise \end{cases} \quad (6)$$

$$x_2, x_3 = \begin{cases} 0.3 & abnormal \\ 0 & otherwise \end{cases}$$

따라서 식(2), (3), (4)에 기반하여 최종 위험도 점수를 산출하고 이는 식 (7)과 같다.

$$risk = w_{call} + w_{facial} + w_{bank} \quad risk \in [0,1] \quad (7)$$

4. 결 론

본 연구에선 ATM의 CCTV 영상을 이용하여 영상 기반 보이스피싱 피해 예방 방법을 제안하였다. 특히, 한국인의 특성에 맞는 한국인 표정 데이터를 이용하여 더욱 한국화된 보이스피싱 피해 예방 시스템을 구축하였다. 또한, 개인의 금융 데이터 즉, 마이데이터를 이용하여 영상 뿐만 아니라 정량적 데이터를 활용하여 예방 방법을 더욱 구체화 시켰다. 또한, 본 연구에서 제안한 모델은 1프레임당 0.01초로 매우 빠른 처리 속도를 가진다.

하지만 본 연구에선 한국인 표정 인식에 대해서 다중 분류를 실시할 때 정확도가 현저하게 낮아지는 한계점이 존재하였다. 또한, 손 탐지에 대한 AP가 낮은 한계점이 존재하였다.

표정의 한계점을 극복하기 위해 서양의 표정 데이터들이 학습된 모델에 전이학습을 하여 더욱 모델의 고도화를 할 계획이다. 또한, COCO-Hand 데이터의 특성상 작은 손의 객체가 있었기 때문에 AP 성능이 낮게 나왔으므로 비교적 객체의 크기가 큰 객체에 대해 추가로 학습하면 더욱 좋은 성능이 나올 것이라 기대한다.

References

[1] https://github.com/ultralytics/yolov5

[2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3] Wang, Chien-Yao, et al. "CSPNet: A new backbone that can enhance learning capability of CNN." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020.

[4] Park, Jongchan, et al. "Bam: Bottleneck attention module." arXiv preprint arXiv:1807.06514 (2018).

[5] Elfwing, Stefan, Eiji Uchibe, and Kenji Doya. "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning." Neural Networks 107 (2018): 3-11.

[6] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.

[7] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International Conference on Machine Learning. PMLR, 2019.

[8] https://www.kaggle.com/andrewmvd/face-mask-detection

[9] Narasimhaswamy, Supreeth, et al. "Contextual attention for hand detection in the wild." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[10] https://aihub.or.kr/aidata/27716

[11] Loshchilov, Ilya, and Frank Hutter. "Decoupled weight decay regularization." arXiv preprint arXiv:1711.05101 (2017).

[12] Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts." arXiv preprint arXiv:1608.03983 (2016).