https://doi.org/10.33778/kcsa.2022.22.5.011

개인정보보호를 위한 다중 유형 객체 탐지 기반 비식별화 기법★

길 예 슬*, 이 효 진**, 류 정 화***, 이 일 구****

요 약

인터넷과 웹 기술이 모바일 장치 중심으로 발전하면서 이미지 데이터는 사람, 텍스트, 공간 등 다양한 유형의 민감 정보를 담고 있다. 이러한 특성과 더불어 SNS 사용이 증가하면서 온라인 상의 개인정보가 노출되고 악용되는 피해 규모가 커지고 있다. 그러나 개인정보보호를 위한 다중 유형 객체 탐지 기반의 비식별화 기술에 관한 연구는 미흡한 상황이다. 이에 본 논문은 기존의 단일 유형 객체 탐지 모델을 병렬적으로 이용하여 다중 유형의 객체를 탐지 및 비식별화 하는 인공지능 모델을 제안한다. Cutmix 기법을 통해 사람과 텍스트 객체가 함께 존재하는 이미지를 생성하여 학습 데이터로 구성하고, 사람과 텍스트라는 다른 특징을 가진 객체에 대한 탐지 및 비식별화를 수행하였다. 제안하는 모델은 두 가지 객체가 동시에 존재할 때 0.724의 precision과 0.745의 mAP@.5 를 달성한다. 또한, 비식별화 수행 후 전체 객체에 대해 mAP@.5 가 0.224로, 0.4 이상의 감소폭을 보였다.

Multi-type object detection-based de-identification technique for personal information protection

Ye-Seul Kil*, Hyo-Jin Lee**, Jung-Hwa Ryu***, Il-Gu Lee****

ABSTRACT

As the Internet and web technology develop around mobile devices, image data contains various types of sensitive information such as people, text, and space. In addition to these characteristics, as the use of SNS increases, the amount of damage caused by exposure and abuse of personal information online is increasing. However, research on de-identification technology based on multi-type object detection for personal information protection is insufficient. Therefore, this paper proposes an artificial intelligence model that detects and de-identifies multiple types of objects using existing single-type object detection models in parallel. Through cutmix, an image in which person and text objects exist together are created and composed of training data, and detection and de-identification of objects with different characteristics of person and text was performed. The proposed model achieves a precision of 0.724 and mAP@.5 of 0.745 when two objects are present at the same time. In addition, after de-identification, mAP@.5 was 0.224 for all objects, showing a decrease of 0.4 or more.

Key words : Image Privacy, De-Identification, Artificial Intelligence, Multi Object Detection, Cutmix

접수일(2022년 10월 01일), 게재확정일(2022년 11월 01일)

★ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2020R1F1A1061107)과 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0008703, 2022년 산업혁신인재성장지원사업), 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0 사업의 연구결과로 수행되었음 (IITP-2022-RS-2022-00156310).

* 성신여자대학교 미래융합기술공학과(주저자)
** 성신여자대학교 미래융합기술공학과(공동저자)
*** 성신여자대학교 융합보안공학과(공동저자)
**** 성신여자대학교 미래융합기술공학과(교신저자)

1. 서 론

사용자들이 자유롭게 데이터를 생성하고 공유하는 SNS(Social Network Service)는 개인을 특정할 수 있는 민감한 정보를 담고 있으며, 동영상 서비스를 기반으로 하는 1 인 미디어는 불특정 다수의 얼굴 등 개인정보가 자신의 의사에 상관없이 노출될 수 있다. 특히, 이미지 데이터는 사람, 텍스트, 공간 등 다양한 유형의 민감 정보를 담고 있다. 이러한 특성과 더불어 SNS 사용이 기하급수적으로 증가하면서 온라인상의 개인정보를 악용한 정보 유출, 사생활 침해, 사칭, 도용 등의 사이버 범죄 역시 급증하고 있다[1].

그러나 개인정보보호를 위한 다중 유형 객체 탐지 기반의 비식별화 기술에 관한 연구는 부족한 상황이다. 다양한 유형의 데이터에 대하여 종래 단일 객체 중심의 비식별화 방식은 개인정보와 민감 데이터를 보호하기 어렵다. 이러한 문제를 해결하기 위해 이미지상에 나타나는 사람과 텍스트 객체를 동시에 식별 및 비식별화하는 기술을 제안한다.

2 장에서는 종래 기법의 한계점을 파악하기 위해 이미지 객체 탐지, 이미지 병합, 비식별화 관련 선행 연구를 분석한다. 3, 4 장에서는 개인정보를 보호하기 위해 다중 유형의 객체를 탐지하고 비식별화하는 모델을 제안하고, 모델의 성능을 입증하기 위해 객체에 대한 탐지율과 비식별화율을 평가한다. 마지막으로 5 장에서는 결론을 제시한다.

2. 배경기술 및 관련연구

2.1. 이미지 객체 탐지

인공신경망 기술의 발전과 함께 이미지 또는 비디오 상의 객체를 식별하는 컴퓨터 비전(computer vision) 기술인 객체 탐지(object detection) 기법이 발전해왔다. 객체 탐지 알고리즘은 한 이미지에서 객체의 위치를 추정하는 지역화(localization)와 객체가 무엇인지 확인하는 분류(classification) 과정을 통해 수행되며, 프로세스 수행 방식에 따라 One-stage 모델과 Two-stage 모델로 나뉜다[2]. 지역화와 분류 작업이 동시에 수행되는 One-stage 모델과 지역화를 수행한 후 분류가 이루어지는 Two-stage 모델은 객체 감지 분야의 두 축을 이루며 발전해왔다. 이러한 수행과정의 차이에 따라 One-stage 모델은 속도에 초점을 두고 있다.

인공신경망 모델인 CNN(Convolutional Neural Networks)과 R-CNN(Region-based Convolutional Neural Networks)기반의 모델은 연구를 거듭하면서 Faster R-CNN 과 같은 모델로 발전하면서 일정 이상의 정확도를 확보하였으나 실시간성이 보장되어야 하는 시스템에 적용할 수 없었다[3].

이후 등장한 YOLO 는 객체 탐지를 수행하기 위해 고안된 심층 신경망으로서 One-stage object detection 방식을 적용하여[4] 기존의 Faster R-CNN 보다 처리속도를 6 배 개선했다. 특히 YOLO v3[5]는 특징이 뚜렷하지 않거나 선명하지 않은 물체에 대해서 Faster R-CNN 이나 SSD(Single Shot Multi-box Detector) 모델 대비 더 높은 인식 정확도와 견고성을 보여주었다[6]. YOLO 는 테두리 상자 조정(bounding box coordinate)과 분류(classification)를 단일 신경망 구조를 통해 동시에 수행한다. 총 세 단계에 거쳐 객체 탐지를 수행하는데, 가장 먼저 이미지를 격자 구조로 분할 하고, 각각 N 개의 초기 탐지 상자와 그에 대한 신뢰도(confidence score)를 부여한다. 신뢰도는 물체를 포함한다는 예측을 얼마나 확신하는지, 물체 정보에 대한 예측이 얼마나 정확할지를 의미하며 수식(1)과 같이 표현할 수 있다.

$$ConfidenceScore : P_r(Object) \times IOU^{truth}_{prediction} \quad (1)$$

수식 (1)에서 $P_r(Object)$는 격자 내에 물체가 존재할 확률을 의미하며, IOU(Intersection over Union)는 두 영역이 겹쳐져 있을 때의 교집합을 합집합으로 나눈 것이다. 즉, 실제 객체의 위치인 truth(Ground-Truth bounding box)와 사용자가 예측한 객체 위치인 prediction(Predicted bounding box)이 중복되는 영역의 크기를 통해

평가한다. 중복되는 영역이 넓을수록 예측 성능이 좋다는 것을 의미하며, YOLO는 이 수치를 기반으로 개별 객체에 대한 검증을 수행한다.

2.2. 텍스트 객체 처리

이미지 내 텍스트 객체 처리는 일반적인 객체 처리와 달리 문자를 검출하기 위한 최소 단위를 정해야 한다. 또한, 이미지 내에서 단어 및 문장 단위로 탐지한 데이터에 대해 맥락을 고려하는 과정이 요구된다. 텍스트 객체 처리 분야에서는 이러한 텍스트의 특성을 반영한 탐지 모델인 OCR(Optical Character Recognition)을 사용한다. OCR은 이미지 내의 글자를 자동으로 인식하는 인공지능 기술로서 입력된 이미지 내 문자를 검출하는 텍스트 탐지(text detection) 과정과 검출된 영역의 문자를 인식하는 텍스트 인식(text recognition) 과정으로 구성된다[7]. OCR 기술은 자동차 번호판 자동인식, 신용카드 광학 인식을 중심으로 연구되었으며 기계학습 기반의 알고리즘과의 결합을 통해 인식률이 향상되고 있다[8]. 그러나 종래의 OCR 기술은 문서에 한정한 기술로 발전했기 때문에 이미지 속의 텍스트 구조를 분석하고 인식하는 데 한계가 존재한다[9].

2.3. 이미지 병합

Cutmix는 Cutout과 Mixup을 합친 데이터 증강 기법의 일종으로, 검정 픽셀 또는 무작위 노이즈 패치를 삽입하는 종래 드롭아웃 방식의 한계점을 개선한 방식이다. Cutmix는 수식 (2)에서 정의된 바와 같이 이미지 x, 라벨 y에 대하여 두 개의 데이터 (x_A, y_A), (x_B, y_B)로부터 새로운 데이터 $(\hat{x}, \hat{y})$를 생성한다[10]. 이때 x의 범위는 2개의 이미지에서 삭제 및 삽입 위치를 나타내는 이진 마스크의 크기인 $W \times H$에 입력 이미지의 채널 크기 C를 곱한 실수 범위 내에서 정의된다. 먼저 두 원본 데이터로부터 잘라낼 영역을 정의하고자 이진 마스크 M을 샘플링 하여 경계 상자 좌표를 설정한다. 이후 자른 면적 비율을 λ로 샘플링 함으로써 자른 면적의 높이와 너비의 비율을 일정하게 유지한다. 합칠 두 개의 이미지를 미니배치 안에서 랜덤하게 선택함으로써 계산상의 오버헤드를 줄인다.

$$\begin{aligned} \hat{x} &= M \odot x_A + (1-M) \odot x_B \\ \hat{y} &= \lambda y_A + (1-\lambda) y_B \end{aligned} \tag{2}$$

해당 기법은 ImageNet-1K 데이터셋에 대하여 다른 데이터 증강 기법 대비 가장 낮은 Top-1 에러율을 보이며, CIFAR-100 데이터셋에 대해서도 가장 낮은 Top-1 에러율을 보인다.

2.4. 비식별화

본 장에서는 선행 연구의 비식별화 방식과 그에 따른 평가지표의 적합성, 그리고 다중 객체에 대한 비식별화 여부를 검토한다.

Safe Fakes는[11] 7가지 익명화 기법을 활용했을 때 도출되는 객체 탐지율을 비교했다. mAP(Mean Average Precision)를 평가지표로 활용함으로써 GAN(Generative Adversarial Network)을 이용한 가명화 방식의 mAP가 원본의 mAP와 유사함을 보였다. 그러나 이 연구는 하나의 이미지상에서 하나의 얼굴 객체에 대한 비식별화 수행을 목표로 함에 따라, 얼굴 이외의 객체의 비식별화를 고려하지 못했다.

DeepPrivacy[12]는 GAN을 활용하여 익명화하고자 하는 대상에 대한 데이터를 생성해낸다. 분류기를 통해 얼굴을 인식한 후, 지정한 얼굴 부분을 잘라내고 생성자를 입력함으로써 익명화된 이미지를 생성했다. 그러나 이 연구는 WIDER-Face 데이터셋을 활용하여 고품질의 이미지를 생성하지만 복잡한 배경정보 및 불규칙한 자세의 경우에는 비현실적인 이미지를 생성하는 한계점이 있다. 또한, 얼굴 이미지를 중심으로 익명화를 수행함에 따라 다른 유형의 객체에 대한 익명화는 고려하지 못했다.

Privacy Intelligence[13]는 온라인 소셜 네트워크상에서의 이미지 공유 수명 주기를 기반으로 하는

개인정보 분석 프레임워크와 함께 개인정보의 속성 및 분류 기준을 제시했다. 이 연구는 안전성과 유용성 간의 trade-off 관계를 고려하면서 보안성 관점에서의 데이터 처리 기준에 기반한 이미지 난독화 방법론을 제시했다. 그러나 방법론만 제시할 뿐, 실제 실험 결과에 대한 평가 및 평가지표를 제시하지 못했다.

Impact of Anonymization[14]은 차량 감지 관련 데이터 보호법을 준수하기 위한 목적으로 개인정보 제거 및 익명화 과정을 수행했으며 이미지 내 객체의 유형 및 크기, 점유율 및 비식별화 방법에 따른 익명화의 영향을 고려했다. 그러나 이 연구는 번호판으로만 객체를 한정하여 비식별화를 수행함에 따라 다른 유형의 텍스트 객체 비식별화가 불가능하다.

텍스트 비식별화 관련 연구인 Preservative[15]는 이미지 품질의 손상을 최소화하면서 자동차 번호판을 식별 및 비식별화하는 IPCB(Inhomogeneous Principal Component Blur) 방법론을 제안했다. 그러나 이 연구는 번호판을 탐지하는 과정을 포함하고 있지 않기 때문에, 비식별화를 적용할 대상의 위치를 정확히 탐지하지 못하는 한계점이 존재한다.

GAN-based[16]는 심층 신경망 기술을 이용하여 GAN 과 DP(Differential Privacy) 기반의 새로운 영상 개인정보보호 방법을 제안했다. 다양한 평가지표(Dhash, SSIM, L0, L2, AlDp) 간 비교 분석을 통해 비식별화 기법에 따른 비식별화 정도를 잘 보여줄 수 있는 평가지표를 선정했다. 그러나 비식별화 기법 적용 시, 얼굴 객체를 중심으로 비식별화를 수행할 뿐 다중 객체에 대한 비식별화는 수행하지 못했다.

<표 1> 비식별화 비식별화 관련 선행 연구 비교표

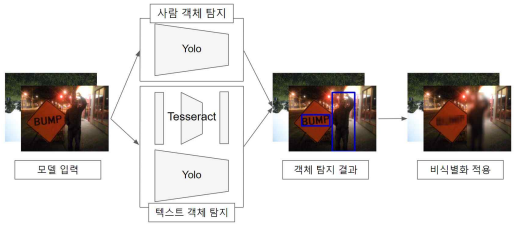
선행 연구	비식별화 모델	평가지표	비식별화 객체
Safe Fakes[11]	GAN	mAP/ FID, L1, L2, LPIPS, SSIM	단일 이미지(얼굴)
DeepPrivacy[12]	GAN	N/A	복수 이미지(얼굴)
Privacy Intelligence[13]	이미지 난독화 기법 (마스킹, 픽셀화, 블러링, 추상화 등)	N/A	단일 이미지
Impact of Anonymization[14]	이미지 난독화 기법 (가우시안 블러, 픽셀화, 픽셀 셔플링, 랜덤값 왜곡 등)	AP	단일 이미지(번호판)
Preservative[15]	IPCB	SSIM	단일 텍스트(번호판)
GAN-based[16]	GAN	Dhash, SSIM, L0, L2, AlDp	단일 이미지(얼굴)

표 1 은 비식별화 관련 선행 연구를 비교한 표이다. 이미지 내 개인정보는 사람의 얼굴과 같은 형태의 정보뿐만 아니라 전화번호, 주민등록번호, 차량 번호판 등과 같은 텍스트 형태의 정보가 함께 존재한다. 그러나 선행 연구들은 사람이나 동물, 번호판이나 표지판 등 동일 유형의 객체에 한정된 비식별화를 수행하기 때문에 개인정보와 관련된 객체에 대한 비식별화를 종합적으로 평가하기 어렵다. 그러므로 다양한 유형의 개인정보 관련 객체를 식별하고 비식별화할 수 있는 기술이 요구된다.

3. 다중 객체 탐지 및 비식별화 모델

제안하는 시스템은 그림 1 과 같이 이미지를 입력으로 받아 객체가 비식별화된 이미지를 결과로 도출한다. 입력으로 주어진 이미지에서 개인정보를 담고 있는 사람과 텍스트 객체를 식별하여 객체를

탐지한다. 이때, 탐지된 객체에 대한 위치 정보인 테두리 상자 정보를 얻을 수 있으며, 이후 이를 이용해 비식별화하는 과정으로 이어진다.

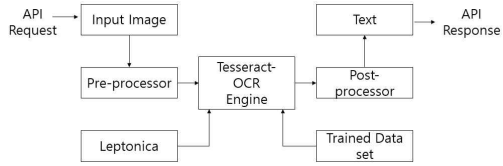


(그림 1) 전체 시스템

3.1. 객체 탐지 모델

객체 탐지 모델은 사람과 텍스트 객체에 대한 탐지 과정을 거친다. 먼저, 사람 객체를 탐지하는 모델은 YOLO v3 를 활용한다. YOLO 모델은 다중 클래스 분류에 대한 구조적 안정성과 탐지율을 개선하기 위해 다양한 기법을 추가하는 등의 과정을 통해 v1 에서 v6 까지 발전하였다. 이 중 YOLO v3 는 실시간 객체 탐지 모델 중 학습과 추론을 모두 고려했을 때 가장 좋은 성능을 보인다[17].

기존의 YOLO 는 텍스트 객체를 탐지하지 않기 때문에 별도로 텍스트 객체를 탐지하는 모델로 학습시키는 과정이 필요하다. 따라서 텍스트 객체 탐지 모델은 YOLO v3 와 더불어 Tesseract 라이브러리를 활용한다. Tesseract 는 다양한 운영체제에서 사용할 수 있는 오픈소스 기반 OCR 라이브러리이다[18]. 전체적인 Tesseract 라이브러리의 프로세스는 그림 2 와 같다. Tesseract 는 입력된 이미지의 특징점을 추출하고 그 특징점을 사용하여 문자를 인식한다. 원본 객체에 대한 테두리 상자(bounding box)를 생성한 후 텍스트가 위치한 방향성에 따라 다각형에 근접하게 특징점을 추출한다. 이후, Tesseract 데이터베이스 검색을 통하여 특징점이 비슷한 문자들과의 Template Matching 을 통해 특징점과 원본 이미지의 오차율이 가장 낮은 문자를 선택하여 텍스트를 인식한다.



(그림 2) Tesseract 객체 탐지 프로세스

따라서 YOLO v3 과 Tesseract 를 함께 이용하여 종래 모델의 한계점을 극복하고 사람 및 텍스트 객체를 포함하는 다중 객체 탐지 및 비식별화 모델을 제안한다.

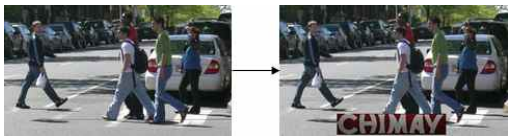
3.2. 비식별화 모델

객체 탐지 모델에서 검출된 객체의 테두리 상자 정보를 바탕으로 OpenCV 의 blur 함수를 사용하여 평균 블러링을 적용한다. 객체 탐지 및 비식별화 과정이 적용된 결과는 그림 3 과 같다. 이렇게 비식별화가 적용된 이미지를 탐지 모델에 입력으로 제공하여 비식별화 기법 적용 후의 객체 탐지율인 비식별화율을 측정한다.

(그림 3) 객체 탐지 및 비식별화 과정

3.3. 데이터 셋

이미지 내 개인정보와 관련된 객체를 사람과 텍스트 두 가지로 한정하였다. 다중 객체 탐지를 위해 구축된 데이터셋 중 사람과 텍스트가 동시에 식별 가능한 데이터셋은 존재하지 않았다. 이에 본 실험에서는 이미지 혼합 및 삭제(image mixing and deleting) 기법 중에서 Cutmix 를 통해 이미지 데이터를 생성하였다. 기존의 사람 객체만 존재하는 데이터셋과 텍스트 객체만 존재하는 데이터셋에서 각각 임의로 이미지를 선택하여 그림 4 와 같이 사람, 텍스트 객체가 함께 존재하는 이미지를 생성하였다.



(그림 4) Cutmix 기법을 통해 생성한 이미지

실험에 이용한 데이터셋은 표 2 와 같다. COCO[22][23]는 다중 객체 탐지를 위한 데이터셋으로 일부 이미지에서 사람 객체와 간판, 교통 표지판 등의 텍스트 객체가 함께 존재한다. 해당 데이터셋의 텍스트 객체의 크기가 매우 작아 탐지하기에 적합하지 않지만, 본 실험에서는 현실적인 환경에서 객체 탐지 성능을 평가하기 위해 COCO 데이터셋을 함께 이용하였다.

<표 2> 객체 탐지 모델 학습에 이용한 데이터셋

dataset	count	label
Pen-Fudan Database [19]	170	person
Pedestrian[20]	1,339	person
TextOCR[21]	21,778	text
COCO2014[22][23]	17,141	person, text

전체 데이터셋에서 학습(train), 검증(validation), 실험(test)에 이용한 데이터의 수는 7:2:1 의 비율이며, 학습에 이용한 데이터 수는 19,670 장, 학습 중 검증에 이용한 데이터의 수는 5,619 장, 마지막으로 실험에 이용한 데이터의 수는 2,809 장이다.

4. 실험 결과

이미지 내 객체에 대한 탐지율(detection rate)과 비식별화율(de-identification rate)을 평가하는 지표로 정밀도(precision), 재현율(recall), F1-score, mAP 를 이용한다. Precision 은 이미지 내 특정 클래스가 존재하는 것으로 예측한 것 중 실제로 해당 클래스의 객체가 존재하는 비율을 의미하며, Recall 은 특정 객체가 존재하는 이미지 중 모델이 해당 클래스가 이미지 내에 존재한다고 예측한 비율을 의미한다[24]. F1-score 는 이론적으로 반비례하는 Precision 과 Recall 의 특성을 고려해 두 지표의 조화평균을 구한 값이다. mAP 는 객체의 실제 테두리 상자와 예측한 테두리 상자의 유사도를 측정하는 비율인 IoU 를 기반으로 혼동 행렬(confusion matrix)을 구성했을 때, 개별 객체의 Precision 과 Recall 이 이루는 곡선 넓이의 평균값을 의미한다. 따라서 mAP@.5 는 IoU 를 0.5 로 고정했을 때의 mAP 를 의미한다.

탐지율은 객체에 대한 탐지율을 나타내는 지표로 Precision, Recall, F1-score, mAP 값이 클수록 큰 값을 가진다. 비식별화율은 비식별화가 수행된 데이터를 객체 탐지 모델에 입력했을 때의 탐지율을 기반으로 값을 평가하며 Precision, Recall, F1-score, mAP 값이 낮을수록 높은 값을 가진다.

<표 3> 비식별화 기법 적용 전후 객체 탐지 모델의 객체 탐지 성능 비교

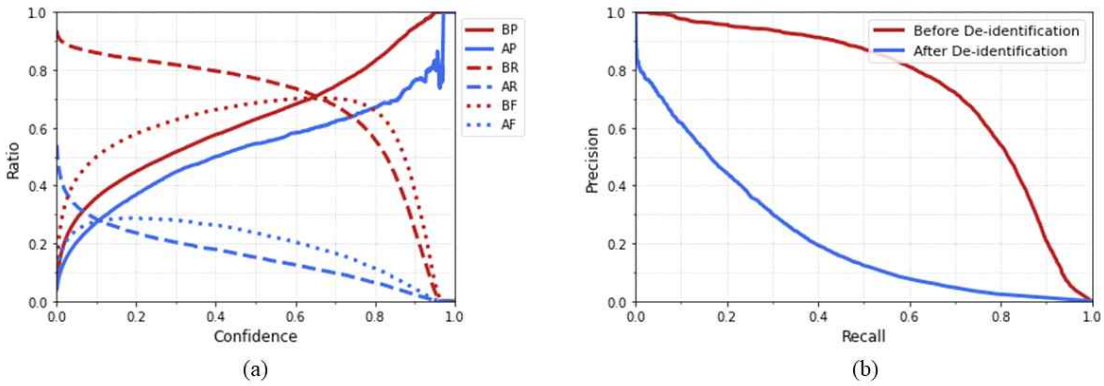
class	precision	recall	mAP@.5	f1-score
all	0.724	0.696	0.745	0.702
person	0.735	0.712	0.773	0.714
text	0.741	0.667	0.716	0.691

(a) Detection Rate - 비식별화 전 성능 지표

class	precision	recall	mAP@.5	f1-score
all	0.373	0.235	0.224	0.288
person	0.492	0.332	0.353	0.396
text	0.236	0.148	0.0958	0.179

(b) De-identification Rate - 비식별화 후 성능 지표

표 3 은 비식별화를 적용하기 전과 후의 탐지율을 비교한 결과를 보여준다. 비식별화를 적용했을 때의 지표 값 표 3 의 (b)가 비식별화 전인 표 3 의 (a)와 비교해 모두 감소했음을 볼 수 있다. 특히 텍스트 객체에 대한 Precision, Recall 값이 약 0.5 정도 감소하면서 사람 객체에 대한 감소폭인 0.3



(그림 5) 비식별화 전, 후의 성능 비교: (a) 신뢰도에 따른 Precision, Recall, F1-score 값의 변화, (b) Recall에 따른 Precision 변화

보다 크게 나타난다. 이에 따라 비식별화 전후의 텍스트 객체의 mAP 감소폭 또한 사람 객체에 대한 mAP 감소폭보다 더 크게 나타났다. 이는 학습에 이용한 텍스트 객체의 크기가 비교적 작고 단순하여 비식별화 기법을 적용했을 때 텍스트 객체의 클래스별 특징에 대한 비식별화 처리율이 더 컸기 때문이다.

그림 5는 비식별화 적용 전후에 따른 평가지표 변화를 나타낸다. 그림 5의 (a)는 신뢰도에 따른 평가지표의 변화를 나타내며, BP(Before De-Identification Precision), BR(Before De-Identification Recall), BF(Before De-Identification F1-score)는 각각 비식별화 전 데이터에 대한 Precision, Recall, F1-score를 의미하고, AP(After De-Identification Precision), AR(After De-Identification Recall), AF(After De-Identification F1-score)는 각각 비식별화 후 Precision, Recall, F1-score를 의미한다. Precision의 경우, 비식별화 기법을 적용했을 때와 적용하지 않았을 때 신뢰도 변화에 큰 영향이 없다. Recall은 비식별화를 적용하지 않았을 때 신뢰도 값이 0.8인 지점까지 일정 이상의 값을 유지하지만, 비식별화를 적용했을 때 0.6보다 작은 값에서 시작하여 일정한 감소폭을 유지한다. 이는 비식별화 모델로 인한 참된 예측의 감소가 Precision보다 Recall에 더 큰 상관관계가 있기 때문이다. 더불어 Recall의 큰 감소는 F1-score 감소로 이어져 비식별화 전에는 약 0.7의 값을 보였지만, 비식별화 후에는 약 0.3의 값을 갖는 것에 그쳤다. 이를 통해 비식별화가 효과적으로 수행되었음을 알 수 있다.

비식별화 전후의 Recall에 따른 Precision인 그림 5의 (b)는 비식별화 전과 비식별화 후의 mAP로 계산되는 면적을 시각적으로 보여준다. 이를 통해 비식별화 후 mAP의 값이 현저히 작아졌음을 알 수 있는데, 이는 비식별화에 의한 Recall이 매우 감소했기 때문이다. 이뿐만 아니라 (b)의 Recall에 따른 Precision 변화 곡선의 기울기나 변곡점이 비식별화 전후로 큰 위치 차이를 보인다. 이를 통해 Precision, Recall, F1-score, mAP의 평가지표가 비식별화 전후로 확연한 차이가 있음을 시각적으로 보여준다.

5. 결 론

본 연구는 개인정보보호 관점에서 사람과 텍스트 객체를 동시에 식별 및 비식별화하는 인공지능 모델을 제안한다. 서로 다른 특징을 가진 사람과 텍스트 객체에 대하여 다중객체 탐지 및 비식별화가 가능함을 입증했다. 제안하는 모델은 두 가지 객체가 동시에 존재할 때 0.724의 precision과 0.745의 mAP@.5를 달성한다. 또한, 비식별화 수행 후 전체 객체에 대해 mAP@.5가 0.224로, 0.4 이상의 감소폭을 보였다. 향후 연구에서는 개인정보를 포함한 객체를 자동으로 식별하고, 주변 객체에 미치는 영향을 최

소화하는 비식별화 기법을 연구할 계획이다. 또한, 본 연구를 실시간 영상 데이터 및 상용 SNS에 적용하여 이미지 생성하고 공유할 때 비식별화 기능이 내장된 플랫폼을 구축한다면 개인정보를 효율적으로 보호할 수 있을 것으로 기대한다.

참고문헌

[1] Jain, A.K., Sahoo, S.R. & Kaubiyal, J. “Online social networks security and privacy: comprehensive review and analysis”. Complex Intell. Syst. 7, 2157 - 2177 (2021)

[2] M. Carranza-Garcia , J. Torres-Mateo, P. Lara-Benítez and J. García-Gutiérrez, “On the Performance of One-Stage and Two-Stage Object Detectors in Autonomous Vehicles Using Camera Data”, Remote Sens. 2021, vol. 13, no. 1, pp. 89, 2020.

[3] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng and R. Qu., “A Survey of Deep Learning -based Object Detection”, arXiv:1907.09408v2, 2019.

[4] H. Chen, Z. He, B. Shi, and T. Zhong, “Research on Recognition Method of Electrical Components Based on YOLO V3”, IEEE Access, vol. 7, pp. 157818 - 157829, 2019.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real -Time Object Detection”, arXiv:1506.02640v5, 2015.

[6] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement”, arXiv:1804.02767, 2018.

[7] Y. Du, C. Li, R. Guo, X. Yin, W. Liu, J. Zhou, Y. Bai, Z. Yu, Y. Yang, Q. Dang and H. Wang, “PP-OCR: A Practical Ultra Lightweight OCR System”, arXiv:2009.09941, 2020.

[8] Bhujbal, Avinash and Deepak T. Mane. “A Survey On Deep Learning Approaches For Vehicle And Number Plate Detection.” International Journal of Scientific & Technology Research 8 (2019): 1378-1383.

[9] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character Region Awareness for Text Detection”, arXiv:1904.01941v1, 2019.

[10] S. Yun, D. Han, S. Oh, S. Chun, J. Choe and Y. Yoo, “CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features”, arXiv:1905.04899, 2019.

[11] SR. Klomp, M. Rijn, R. Wijnhoven, C. Snoek and P. De With, “Safe Fakes: Evaluating Face Anonymizers for Face Detectors”, Proceedings of 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), 2021.

[12] H. Hukkelas, R. Mester. and F. Lindseth, “DeepPrivacy: A Generative Adversarial Network for Face Anonymization”, arXiv:1909.04538, 2019.

[13] C. Liu, T. Zhu, J. Zhang and W. Zhou, “Privacy Intelligence: A Survey on Image Privacy in Online Social Networks”, arXiv:2008.12199v2, 2020.

[14] L. Schnabel, S. Matzka, M. Stellmacher, M. Patzold and E. Matthes, “Impact of Anonymization on Vehicle Detector Performance” Proceedings of 2019 Second International Conference on Artificial Intelligence for Industries (AI4I), 2019.

[15] L. Du. and H. Ling., “Preservative License Plate De-identification for Privacy Protection” Proceedings of 2011 International Conference on Document Analysis and Recognition, 2011.

[16] J. Yu, H. Xue, B. Liu, Y. Wang, S. Zhu and M. Ding “GAN_based differential private image privacy protection framework” Sensors 2021, vol. 21, no. 1, pp. 58, 2020.

[17] Z. Ge, S. Liu, F. Wang, Z. Li and J. Sun, “YOLOX: Exceeding YOLO Series in 2021”, arXiv:2107.08430, 2021.

[18] Nanonets, “How to OCR with Tesseract, OpenCV and Python”. Nanonets, July 2022, https://nanonets.com/blog/ocr-with-tesseract/, accessed 30 Sep 2022.

[19] L. Wang, J. Shi and G. Song& I. Shen, “Object detection combining recognition and segmentation.”, Asian conference on computer vision, Springer,

Berlin, Heidelberg, 2007.

[20] N. J. Karthika, and S. Chandran, “Addressing the False Positives in Pedestrian Detection.", Electronic Systems and Intelligent Computing, Springer, Singapore, 2020. 1083-1092.

[21] A. Singh, G. Pang, M. Toh, J. Huang, W. Galuba and T. Hassner, “TextOCR: Towards large-scale end-to-end reasoning for arbitrary-shaped scene text.", Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021.

[22] TY. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, L. Zitnick and P. Dollar “Microsoft coco: Common objects in context.", European conference on computer vision, Springer, Cham, 2014.

[23] A. Veit, T. Matera, L. Nemann, J. Matas and S. Belongie, “Coco-text: Dataset and benchmark for text detection and recognition in natural images.", arXiv:1601.07140, 2016.

[24] Zhu, Mu. “Recall, precision and average precision.” Department of Statistics and Actuarial Science, University of Waterloo, Waterloo 2.30 (2004): 6.

〔저자소개〕

길 예 슬 (Ye-Seul Kil)
2021년 8월 성신여자대학교 융합보안공학과 학사
2021년 9월 성신여자대학교 미래융합기술공학과 석사 재학
email : kilyeseul7@gmail.com

류정화 (Jung-Hwa Ryu)
2019년 3월 성신여자대학교 융합보안공학과 학사 재학
email : yoorisu7@gmail.com

이 효 진 (Hyo-Jin Lee)
2022년 2월 성신여자대학교 통계학과, 정보시스템공학과 학사
2022년 3월 성신여자대학교 미래융합기술공학과 석사 재학
email : 220226032@sungshin.ac.kr

이 일 구 (Il-Gu Lee)
2003년 2월 서강대학교 전자공학과 학사
2005년 2월 KAIST 정보통신대학원 석사
2012년 2월 KAIST 지식재산대학원 석사
2016년 2월 KAIST 전산학부 박사
2017년 2월~현재: 성신여자대학교 융합보안공학과/미래융합기술공학과 조교수
email : iglee@sungshin.ac.kr