인공지능 학습용 데이터의 개인정보 비식별화 자동화 도구 개발 연구 - 영상데이터기반 -

[1] 이현주, [2] 이승엽, [3*]전병훈

Research on the development of automated tools to de-identify personal information of data for AI learning - Based on video data -

[1]Hyunju Lee, [2]Seungyeob Lee, [3*]Byunghoon Jeon

요 약

최근 데이터 기반 산업계의 오랜 숙원이었던 개인정보 비식별화가 2020년 8월 데이터3법[1]이 개정되어 명시화 되었다. 4차 산업시대의 원유[2]라 불리는 데이터를 산업 분야에서 활성화할 수 있는 기틀이 되었다. 하지만, 일각에서는 비식별개인정보(personally non-identifiable information)가 정보주체의 기본권 침해를 우려하고 있는 실정이다[3]. 이에 개인정보 비식별화 자동화 도구인 Batch De-Identification Tool을 개발 연구를 수행하였다.

본 연구에서는 첫 번째로, 학습용 데이터 구축을 위해 사람 얼굴(눈, 코, 입) 및 다양한 해상도의 자동차 번호판 등을 라벨링하는 이미지 라벨링 도구를 개발하였다. 두 번째로, 객체 인식 모델을 학습하여 객체 인식 모듈을 실행함으로써 개인정보 비식별화를 수행할 수 있도록 하였다.

본 연구의 결과로 개발된 개인정보 비식별화 자동화 도구는 온라인 서비스를 통해 개인정보 침해 요소를 사전에 제거할 수 있는 가능성을 보여주었다. 이러한 결과는 데이터 기반 산업계에서 개인정보 보호와 활용의 균형을 유지하면서도 데이터의 가치를 극대화할 수 있는 가능성을 제시하고 있다

Abstract

Recently, de-identification of personal information, which has been a long-cherished desire of the data-based industry, was revised and specified in August 2020. It became the foundation for activating data called crude oil[2] in the fourth industrial era in the industrial field. However, some people are concerned about the infringement of the basic rights of the data subject[3]. Accordingly, a development study was conducted on the Batch De-Identification Tool, a personal information de-identification automation tool.

In this study, first, we developed an image labeling tool to label human faces (eyes, nose, mouth) and car license plates of various resolutions to build data for training. Second, an object recognition model was trained to run the object recognition module to perform de-identification of personal information.

The automated personal information de-identification tool developed as a result of this research shows the possibility of proactively eliminating privacy violations through online services. These results suggest possibilities for data-based industries to maximize the value of data while balancing privacy and utilization.

Keywords: *Data for AI training, Auto Labeling, De-identification, Face Recognition, License plate recognition*

[1] *동국대학교, 기술창업학과, 박사수료(hj.lee@maxted.kr)*
[2] *동국대학교, 기술창업학과, 박사수료(s.lee@maxted.kr)*
[3*] *교신저자 동국대학교 기술창업학과 교수(bhjeon@dongguk.edu)*

Received: May. 08, 2023, Revised: Jun. 13, 2023, Accepted: Jun. 21, 2023

I. 서론

세계적으로 4 차 산업혁명에 따라 산업 분야별 경계를 벗어나 빅데이터, 인공지능, 클라우드, 사물인터넷 등의 핵심 기술을 활용하여 기업들이 디지털 트랜스포메이션을 모색하여 제품을 출시하고 서비스되고 있다. 이는 국가 경쟁력 강화를 위한 중요한 산업 Momentum 이라 할 수 있다. 그러나 데이터 중 사이버 공간에서 특정인을 식별할 수 있는 디지털화된 개인정보로부터 사생활 비밀, 자유를 비롯한 개인의 주체성과 존엄성 확보 등 개인정보는 보호되어야 한다. 이러한 양면성을 가진 개인정보 데이터 활용을 위한 방안으로 미국, 일본, 유럽(EU) 등 우리나라를 포함한 세계 각국은 개인정보의 비식별화 정책을 수립하여 실시하고 있다[4].

본 연구에서는 4 차 산업혁명에서 주로 사용되는 인공지능 학습용 데이터 중 정적인 이미지, 동적인 이미지로 제공되는 사람 얼굴(눈, 코, 입), 차량번호 등의 개인정보를 비식별화 작업을 자동으로 수행할 도구를 개발하고, 다양한 해상도의 자동차 번호판 이미지를 통해 인공지능 모델을 수행하는 Batch De-Identification Tool 개발과정을 소개하고자 한다.

본 논문의 구성은 2 장에서는 인공지능 학습 데이터, 개인정보 침해 요소와 개인정보 비식별화 방법 등 연구 배경 및 선행연구를 설명하고, 3 장에서는 이미지 데이터 라벨링부터 비식별화 자동화를 위한 비식별화 서버 구축 및 연구 결과를 설명합니다. 마지막 4 장에서는 활용 방안을 제시한다.

II. 연구배경 및 선행연구

2.1. 인공지능 학습용 데이터

인공지능 학습용 데이터는 머신러닝, 딥러닝 등 인공지능 모델 학습을 위해 사용되는 데이터를 말한다. 인공지능 학습용 데이터는 원본 데이터와 라벨링 데이터로 구성된다. 원본 데이터는 이미지, 영상, 텍스트, 음성 등이고 라벨링 데이터는 활용 목적에 맞게 라벨링 작업이 수행된 데이터이다. 원본 데이터와 라벨링 데이터는 1 대다로 라벨링 데이터는 다양한 형식으로 생성될 수 있다. AI 학습용 데이터의 학습 및 평가 진행 개념도[5]를 나타낸 ‘그림 1’을 보면 인공지능 학습용 데이터는 목적에 따라 학습 데이터, 평가 데이터로 구분되고 학습 데이터 70~80%와 평가 데이터 20~30%로 구분된다. 학습 데이터는 다시 학습 데이터 50%, 검증 데이터 30%로 구분한다. 검증 데이터는 학습을 마친 모델의 예측 및 분류의 정확도를 검증하는데 사용된다. 평가 데이터는 모델이 학습하지 않은 데이터로 모델의 예측 성능을 평가하는데 사용된다.

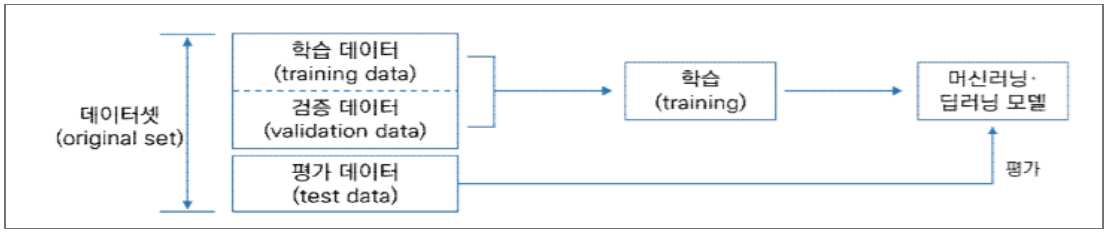

Figure 1. Concept Map of Learning and Assessment Progression for Data for Learning (NIA, "IT & Future Strategy Report")

그림 1. 학습용 데이터의 학습 및 평가진행 개념도(NIA, 「IT & Future Strategy 보고서」)

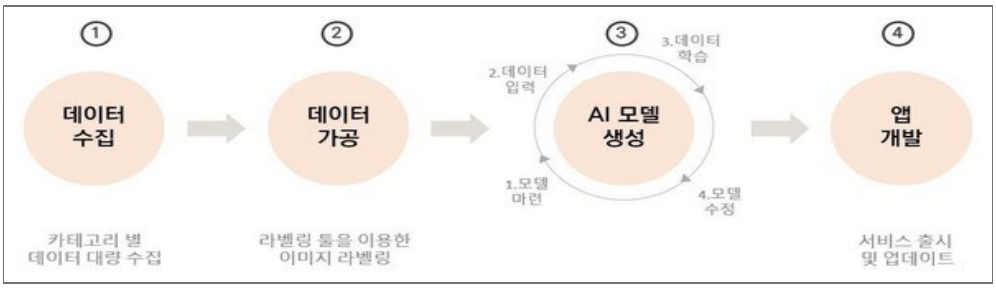

Figure 2.The process of collecting data for AI learning and launching AI services (NIA, "IT & Future Strategy Report")

그림 2.AI 학습용 데이터 수집 및 AI 서비스 출시까지의 과정(NIA, IT & Future Strategy 보고서」)

'그림 2'는 인공지능 학습용 데이터 수집에서 AI 서비스 출시까지의 과정을 도식화한 것이다. 1 번) 데이터 수집은 AI 모델생성을 위한 데이터 수집, 제작하여 데이터셋을 구축한다. 2 번) 데이터 가공, 라벨링 작업으로 이미지, 영상, 텍스트 등의 원본 데이터를 인공지능의 학습기반을 조성하기 위해 편향과 노이즈를 제거하고 원하는 객체(사람, 자동차등)의 위치를 표시한다. 이를 위해 박스나 점을 찍고 객체 라벨링(식별자, 정답표, 속성) 작업을 진행한다. 이러한 데이터 라벨링 작업은 AI 모델생성 전에 이루어지며 대체로 수작업에 의해서 처리되고 인공지능 학습 시간의 약 80~90%를 라벨링 작업이 주를 이루게 된다[6]. 3 번) 전 처리된 데이터로 모델을 학습시켜 학습률(Learning Rate)을 조정하며 정확도(Accuracy)를 높이는 과정을 반복하면 최적의 모델을 생성된다. 4 번) 학습된 최적의 모델을 이용하여 서비스를 제공할 SW 를 출시한다

2.2. 개인정보 비식별화

개인정보는 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보를 의미한다. 비식별화는 개인정보의 일부 또는 전부를 삭제, 대체하거나 다른 정보와 쉽게 결합하지 못하도록 하는 일련의 조치로, 특정 개인을 유추할 수 없도록 한다. 데이터 중 사이버공간에서 특정인을 식별할 수 있는 디지털화된 개인정보로부터 사생활 비밀, 자유, 개인의 주체성, 존엄성 등이 보호되어야 하므로 개인정보를 취급하는 기업이나 조직에서 개인정보를 보호하기 위해 개인정보를 비식별화하여 특정 개인과 연관성을 파악하기 어렵게 만들어야 한다[7].

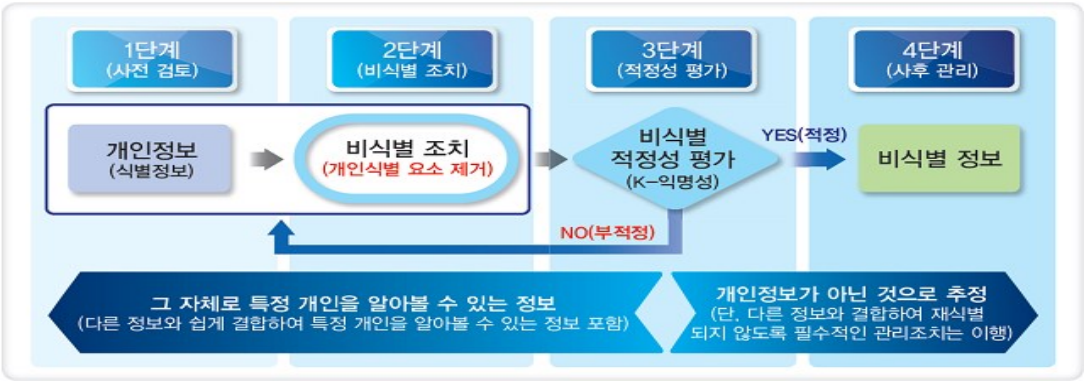

Figure 3. De-identification measures and post-management procedures [7]

그림 3. 비식별 조치 및 사후관리 절차 [7]

'그림 3'의 비식별화 조치 및 사후관리 절차를 보면 1 단계(사전 검토)에서 데이터가 개인정보로 식별될 가능성이 있는지 사전에 검토한다. 만약, 개인정보가 식별될 가능성이 있다면, 2 단계(비식별화 조치)에서 개인정보 비식별화를 위해 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등의 작업을 수행하며, 이를 위해 ISO/IEC 20889, NIST 비식별화 처리 가이드

라인, 익명화 프레임워크, 빅데이터 프라이버시 설계 등 다양한 개인정보 비식별화 표준안에서 제시된 조치법을 활용할 수 있다. 3 단계(적정성 평가)에서는 비식별화된 데이터가 재 식별될 가능성을 최소화하기 위해 k-익명성, l-다양성, t-근접성 등의 프라이버시 보호 모델을 적용한다. k-익명성은 동일한 속성 값을 가진 데이터 집합 내에서 개인을 식별할 수 없도록 최소 k 개 이상의 데이터가 존재하도록 하는 것이고, l-다양성은 다양한 속성 값을 가진 데이터를 동일한 속성 값으로 대체하여 k-익명성을 달성하면서도 정보 손실을 최소화하는 것을 목적으로 하는 기법이고, t-근접성은 비식별화된 데이터가 원래 개인정보 데이터와 충분히 유사하도록 유지하는 것이 목적이고 t-근접성을 달성하면 개인정보 비식별화 기법이 충분히 안전하며, 재식별화의 위험이 감소하게 된다. 4 단계에서는 재식별 방지 및 비식별화 정보에 대한 안전성 조치로 사후관리를 실시한다.

2.3. 영상 데이터 비식별화 기술

네비게이션의 자료로 사용되는 거리 지도, 스마트 기기에 의한 동영상, 홍보 게시물, 블랙박스와 같은 인공지능 학습 데이터로 사용되는 영상 데이터는 다양한 영역에서 개인의 프라이버시 침해 요소를 대비하기 위해 k-익명화 기술이 필요하다. 이를 위해 2 단계로 구성된 영상 데이터 익명화 기술은 식별 가능한 개인의 영역을 탐지하는 기술과 탐지된 식별 영역을 변형하는 기술로 구성된다.

한국정보화진흥원에서 발간한 "영상 데이터 익명화 기술 및 평가방안"[8]에 따르면, 영상 데이터는 거리 지도, 스마트 기기에 의한 동영상, 홍보 게시물, 블랙박스 등에서 인공지능 학습 데이터로 사용되며, 이러한 데이터는 k-익명화 기술을 적용하여 개인정보 보호를 강화해야 한다. 해당 보고서에서는 이미지 필터링, 이미지 암호화, 얼굴 합성, 인페이팅과 같은 비식별화 기술이 제시되고 있다.

첫째, 이미지 필터링은 영상의 각 프레임에 여러 개의 필터들을 적용하여 영상을 변형하는 기법으로 개인정보 식별이 어렵지만, 딥러닝 기술의 발전으로 필터링 제거하여 원본 이미지로 복원할 수 있는 단점이 있다. 이미지 필터링 종류에는 가우시안 함수를 사용하여 이미지를 흐리게 만들고 상세한 특징을 제거하는 방법인 블러(blur)와 이미지 픽셀 수를 줄여 해상도를 낮추는 방법인 픽셀레이트(pixelate)가 있다.

둘째, 이미지 암호화는 영상의 일부분 또는 전체를 암호화하여 가명 처리하는 기술로 복호화 과정을 거쳐 복원할 수 있지만, 연산양이 많아 실시간 처리에는 적합하지 않다. 이미지 암호화 기술로는 주파수 영역으로 변환한 영상의 일부분만 암호화하는 DCT(Discrete Cosine Transform) 기반 기술과 픽셀의 위치를 규칙에 따라 변화시키는 픽셀 위치 기반 기술이 있다.

셋째, 얼굴 합성 기술은 k-익명성 모델을 확장한 k-Same 모델 함수를 사용하여 주어진 얼굴 이미지의 집합에 대해 k 개의 얼굴 이미지의 평균으로 합성된 새로운 이미지를 생성하여 대체하는 방식으로, 1/k 이 넘지 않게 해야 한다.

네 번째, 인페이팅 기술은 특정 부분을 제거하고 다른 이미지로 채우는 기법으로, 주어진 영상 내에서 공백과 가장 비슷한 영역을 찾아 채우는 방식의 패치 기반 인페이팅과 배경과 객체로 구분하여 제거하고 남은 공백을 배경으로 대체하는 방법의 객체 기반 인페이팅이 있다.

2.4. 개인정보보호 강화

개인정보보호위원회의 보도자료[9]에 따르면 AI 허브를 통해 구축된 인공지능(AI) 학습용 데이터 구축 단계에서 인터넷진흥원을 통해 표본 샘플 63 종에 대해 개인정보 포함 여부 및 재식별 가능성 등을 검토하여 과기정통부와 한국지능정보사회진흥원은 비식별화 조치를 위하여 개인정보가 포함되어 있을 가능성이 있는 65 종, 1 억 8 천여건의 인공지능(AI) 학습용 데이터를 최종 점검하고 가명 처리 등 개인정보 비식별화를 추가 진행하였다고 한다. 그중 이미지, 영상 데이터 46 종, 8323 만건은 얼굴, 차량번호 등을 식별할 수 있는 경우 마스킹 작업 처리했으며 나머지 19 종, 9970 만건은 비정형 문자 데이터로 이름, 전화번호, 주소 등 개인정보가 확인되는 경우 보통명사와 기호의 조합으로 가명처리해 제공하였다. 이와 같이 인공지능 학습용 데이터를 수집하고 라벨링하여 데이터셋을 구축하는 과정에서 개인의정보가 유출되지 않도록 사전에 개인정보 비식별화 자동화 도구의 필요성이 절실한 실정이다

2.5. Deep Face recognition

얼굴 인식은 이미지에서 하나 이상의 얼굴을 식별하거나 확인하는 작업이다. '그림 4'와 같은 얼굴 감지 시스템은 이미지에서 얼굴 추출(face detection) 하기 위해서는 일련의 얼굴 이미지 전처리 작업이 필요하다. 우선 사진이나 영상자료에서 사람의 얼굴만 있는 것이 아니므로 얼굴 부분만 cropping 하여 Face Alignment(눈, 코, 입 등 얼굴이라는 특징점 검출) 즉 face landmark detection 하여 얼굴의 위치를 정면을 향하도록 조절한 후 Face processing 과정을 거쳐 Image Augmentation(이미지 증식)된 데이터셋을 사용하여 모델 훈련과정[10]에서 Neural Network 을 통해 각 얼굴의 feature 를 추출하고 loss 가 적은 방향으로 학습한다.

손실 함수(또는 비용함수)는 모델을 학습할 때 오류를 최소화하여 성능이 가장 잘 나오게 하는 함수를 의미한다. 손실 값이 최소화되도록 하는 가중치(weight)와 편향(bias)를 찾는 방법을 Optimization(최적화), Generalization(일반화)라고 할 수 있다. 비용, 손실이 얼마나 발생했는지 나타내는 함수가 Cost Function(비용함수), Loss Function(손실함수)이라 한다

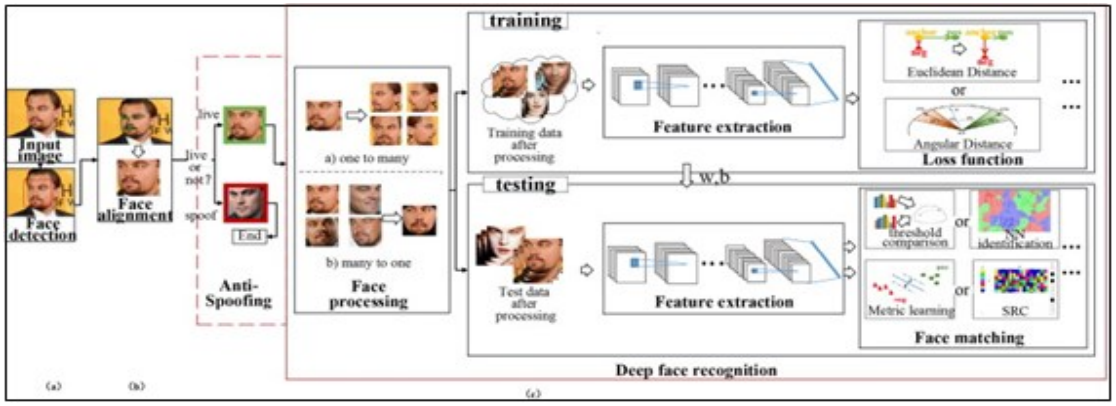

Figure 4. Deep face recognition system with face detector and alignment function [10]
그림 4. 얼굴 감지기 및 정렬 기능이 있는 Deep Face recognition 시스템 [10]

보통 얼굴인식에 쓰이는 네트워크는 ICPvR(International Conference on Plasmodium vivax Research)에서 발표된 것들을 많이 참고하는데 그중 ResNet 을 많이 선호하고 휴대폰을 중점 타겟으로 하는 기업과 고객을 위해서는 MobileNet 을 많이 선호하는 편이다[25] 얼굴인식분야 CNN 모델 훈련에 필요한 손실함수를 알맞게 선택하여 정확도를 높이고 있다. [그림 5] LossFunction 발전과정을 보면 Losss 개발에는 Mertric Learning 에 Contrastive Loss, Triplet Loss, Margin Based Classification 에 해당하는 Softmax with CenterLoss, Sphere Face, Soft-Margin Loss, AM-Softmax, ArcFace 등이 포함된다.[10] 각 단계 중 먼저 Softmax Loss 의 소프트맥스함수는 출력 값이 모두 양수로 그 값의 합이 1 이 되고 함수의 결과값 분포와 정답 라벨 분포와의 유사도를 cross entropy 로 측정함으로써 손실함수를 계산한다. Softmax Loss 는 다중 클래스 분류 문제에서 사용되는 손실 함수 중 하나로 소프트맥스 함수를 적용한 예측값과 실제값의 분포를 비교하여 손실을 계산한다. 2014 년도 DeepFace, DeepID 계열에서 소프트맥스 손실함수가 활용된 것을 볼 수 있다. 세 번째에 위치하는 "Triplet loss" 는 학습 데이터 쌍을 3 개를 이용하는 방법으로 2 개는 동일 클래스에 속하도록 한다. 3 개의 데이터를 선택하는 부분이 복잡하지만, 성능이 좋아 많은 메트릭 러닝에서 사용되었다. '그림 5'에서 마지막에 위치한 "Large margin loss" 는 임베딩 벡터와 클래스 중심 간 거리의 손실값 계산 시 정답 클래스 중심까지의 거리 계산에서 margin 값을 더해주며 모델을 훈련시키는 방법(Large margin loss)으로 이를 통해 feature vector 가 메트릭 러닝이 필요로 하는 동일 클래스 내에서는 더 잘 모여 있고, 다른 클래스와는 더 멀리 떨어지는 곳으로 임베디드 되었다.

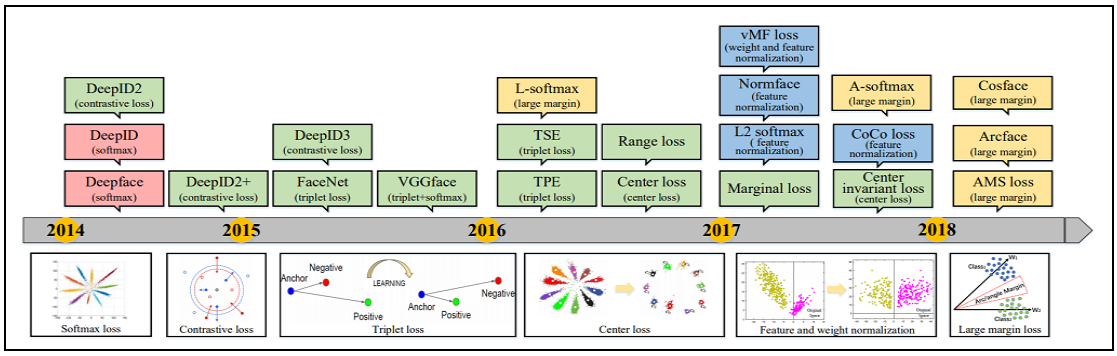

Figure 5. Loss Function Evolution.[10]
그림 5.Loss Function 발전과정[10]

ArcFace 는 Large margin loss 를 활용하는 것으로 유클리드 거리를 사용하는 대신 하이퍼스피어에서 측지 거리라고 하는 점 사이의 최단 거리를 나타낸다. '그림 6'은 ArcFace 의 기하학적 해석을 나타낸 것으로[13] (a) 파란색과 녹색 점은 서로 다른 두 클래스의 내장형 특징을 나타내고 ArcFace 는 클래스 간에 직접 각도(arc)의 여백을 부과할 수 있다. (b) 각도와 호 여백 사이의 대응성을 보여주며 ArcFace 의 각도 여백은 하이퍼스피어 표면의 호 여백(지형 거리)에 해당한다.

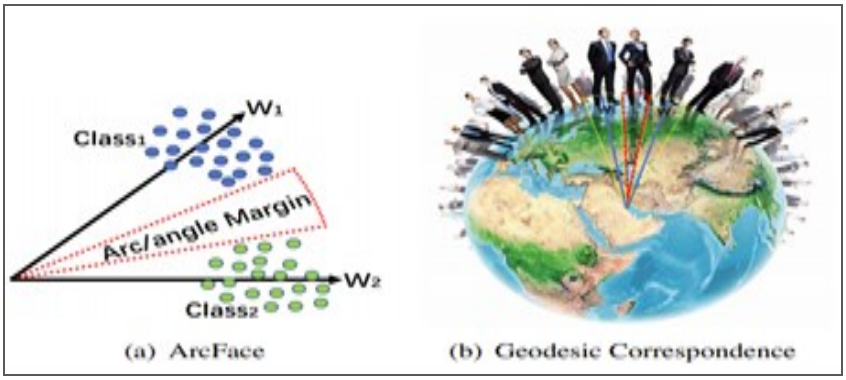

Figure 6.Geometric interpretation of ArcFace
그림 6.ArcFace 의 기하학적 해석.

ArcFace 손실은 SoftMax 의 로짓을 변경하며 하이퍼스피어의 측지 거리와 정확히 일치하기 때문에 기하학적 해석이 명확하다. ArcFace 손실은 마진, 즉 정규화된 가중치 및 기능 덕분에 하이퍼스피어의 측지 공간에서 결정 경계를 최대화한다. 이는 얼굴 인식을 위한 매우 구별되는 기능을 얻고 무시할 수 있는 계산 오버헤드로 쉽게 구현할 수 있는 것이다.

III. 비식별화 자동화 도구 개발 연구

3.1 개발 환경

인공지능 데이터 개인정보의 비식별화 자동화 도구 개발 환경은 다음과 같다. OS 로 Ubuntu 16.0.4(TLS) Server, 개발 언어로 Pyton 3.6.1, GPU 사용과 Computer Vision, Object Dectection 을 위해서 CUDA 10.1, CUDNN 7.5.1, Nvidia-drivere 418.67, OpenCV 4.4, GCC 5.4.0, Cmake3.5.1 를 사용하였다.

3.2. 개발 설계 구성

본 연구의 전체 설계도는 '그림 7'과 같이 웹 플랫폼을 통한 작업지시, 전처리, 작업수행까지의 플로우를 담고 있다. 첫 번째 작업지시는 웹 플랫폼을 통해 작업자가 영상 업로드, 학습데이터 업로드하여 학습 진행 등 작업을 지시한다. 두번째, 전처리과정은 웹 플랫폼에서 수동 라벨링 작업과 데이터를 취합한다. 마지막으로 작업수행단계에서 비식별화 서버는 작업지시에 따라 다중 GPU 를 사용하여 데이터 학습을 진행하여 영상 속 객체(얼굴, 번호판)를 검출하고 비식별화 프로세스를 거쳐 데이터를 저장하는 작업을 수행한다.

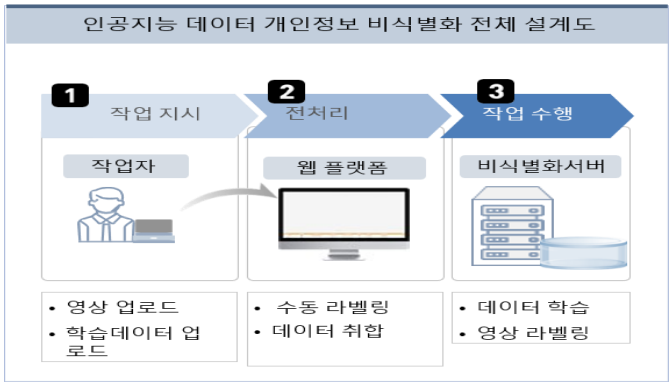

Figure 7. Overall project blueprint
그림 7.프로젝트 전체 설계도

'표 1'는 각 패키지 구조를 나타낸 것이고, '그림 8'은 실제 구축된 프로젝트 패키지다.

Table 1.Overall project blueprint
표 1. 프로젝트 패키지 구조

Package Name	Description	Package Name	Description
Makefile	Files used to build the project	trainWeight	Weighted files used for training
darknet*	Generated upon completion of project build	backup	Saving weighted files generated during training
3rdparty	Libraries needed for de-identification	data	Ground truth files used for detection
dataset	Training Data	input_video	Labeling and detection images directory
cfg	Setting up training and detection options	output_image	Image labeling coordinates storage directory
chart.png	Training result chart image	output_video	Labeled and de-identified images storage directory
weights	Weighted file directory	output_video_label	Image label coordinates storage directory

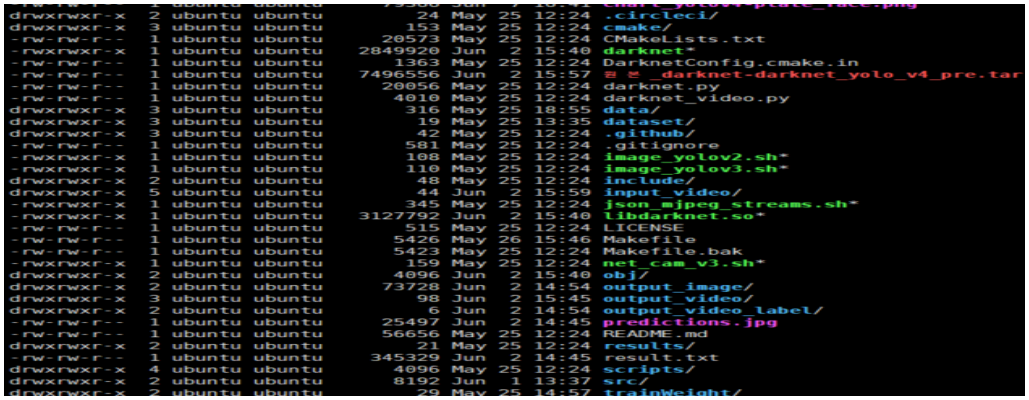

Figure 8.Project Package Structure
그림 8.프로젝트 패키지 구조

3.3. 이미지 데이터 라벨링

데이터 라벨링은 인공지능 학습에 사용할 데이터들을 바운드 박스나 점선을 찍어 검출하고 바운딩된 영역에 태그 즉 라벨을 붙이는 것을 말한다[6]. 본 연구에서는 사람 얼굴 (눈, 코, 입), 자동차 번호판 등 민감한 개인정보를 비식별화 자동화 개발을 위한 인공지능 학습용 데이터 구축을 위한 이미지 데이터 라벨링 작업을 웹기반에서 여러 명이 동시에 수행할 수 있는 Max Data Platform[12]을 사용하였다. '그림 9'는 인공지능 학습데이터를 구성한 것이다. 학습데이터는 총 3 개의 파일, 학습에 사용될 이미지, 해당 이미지의 라벨 좌표, 이미지의 경로가 입력된 파일로 구성된다. '그림 10'은 이미지 경로구성 파일이며, '그림 11'는 이미지의 좌표 파일 구성본이다.

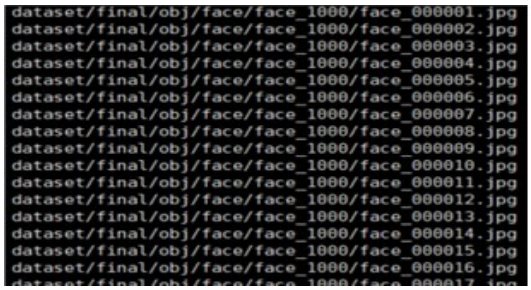

Figure 9.Organizing AI training data
그림 9.인공지능 학습데이터 구성

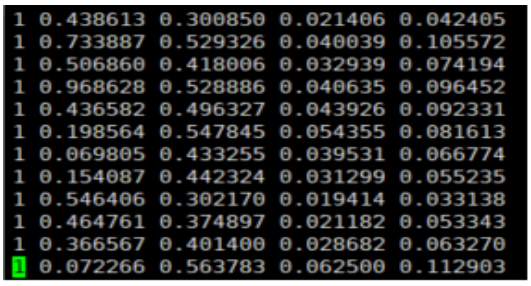

Figure 10.Configuring Image Pathfiles
그림 10.이미지 경로파일 구성

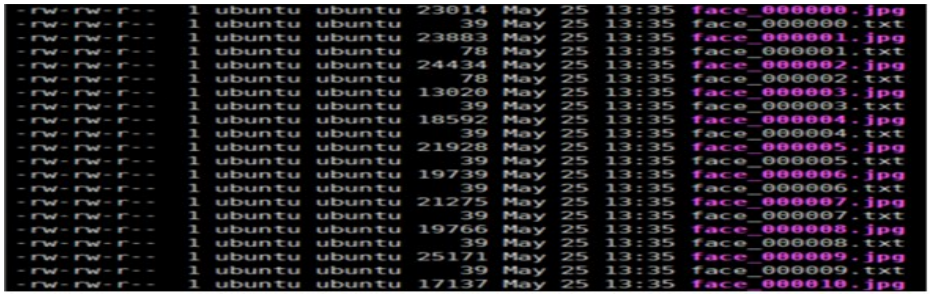

Figure 11.Organizing Coordinate Files
그림 11.좌표 파일 구성

3.4. 객체 인식 모델 개발

ArcFace 손실함수를 사용하여 대규모 얼굴 인식 모델을 위해 DCNN(Deep Convolutional Neural Network)의 [11] 알고리즘을 통해 학습된 Feature Embeddings 의 분별력을 효과적으로 높여 얼굴 인식 기능을 얻으려 한다. 얼굴 인식에 대한 매우 분별력 있는 Features 를 얻기 위한 ArcFace(Additive Angular Margin Loss, 추가각도 여유손실)을 이용하여 기초 교육용 데이터셋과 우리가 직접 구축한 데이터셋을 활용하여 머신러닝을 수행한다. ArcFace 는 이전 선행연구와는 다르게 안정적인 성능을 가지기 위해서 다른 loss function 과 결합될 필요가 없으며, 어떤 training dataset 에도 쉽게 수렴할 수 있다 [11]. 임베딩 생성 기능 추출기로 사용하기 위해 LResNet100E-IR 로 ArcFace 교육하였다 이 모델은 잔여 네트워크로 ArcFace 손실로 학습되어 훈련하는 동안 모든 이미지는 RetinaFace 검출기를 사용하여 자르고 정렬된다[13] . ArcFace 와 유사한 결과를 가지는 CosineFace 도 마진 개선 소프트맥스 방법의 우수성은 트리플트 손실 미세 조정을 통해 유지되므로 LResNet50E-IR 로 CosineFace 교육하고 LMobileNet-GAP 로 소프트맥스 학습 후 소프트맥스 모델을 트리플트 손실과 함께 학습하였다.

'그림 12' 부터 '그림 17'까지는 학습 결과를 차트로 시각화한 것이며, 학습이 진행됨에 따라 손실률(avg loss)이 감소하여 정상적으로 학습이 진행됨을 알 수 있다. 또한 학습데이터가 증가함에 따라 mAP(알고리즘 성능 평균), avg loss 값이 감소하고 mAP 감소가 멈추는 시점으로 보아 충분한 학습이 완료되었음을 알 수 있다

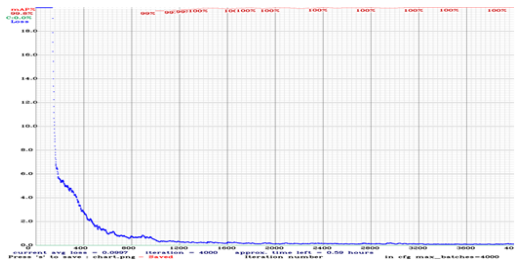

Figure 12. Learning 1st
그림 12.학습 1 회차

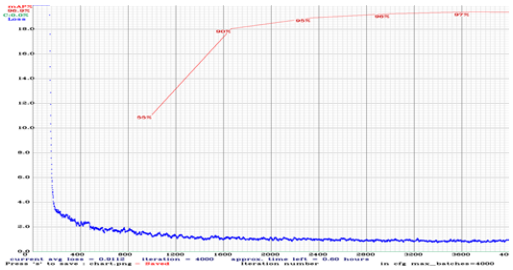

Figure 13. Learning 2nd
그림 13.학습 2 회차

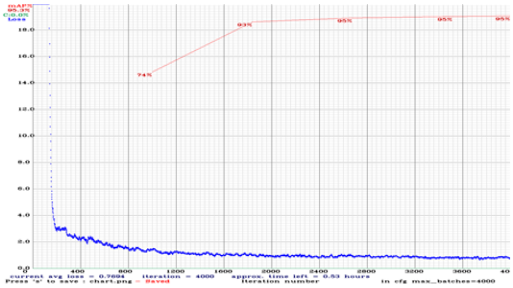

Figure 14. Learning 3rd
그림 14.학습 3 회차

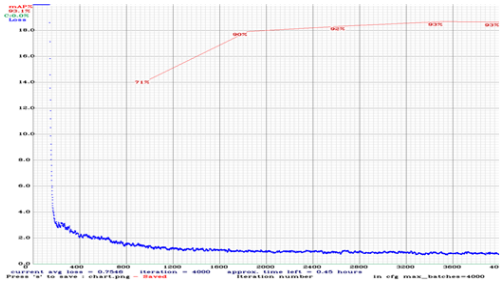

Figure 15. Learning 4th
그림 15.학습 4 회차

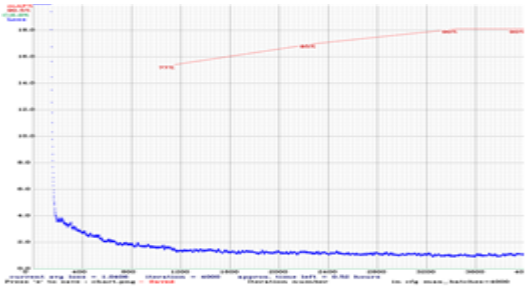

Figure 16. Learning 5th
그림 16.학습 5 회차

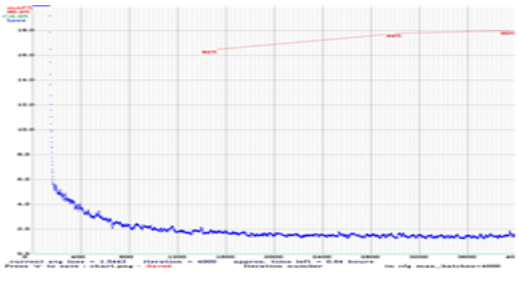

Figure 17.Learning 6th
그림 17.학습 6 회차

3.5. 객체 인식 모듈 실행

객체인식모델을 이용하여 학습이 이루어졌으므로 테스트 데이터셋을 활용하여 엔진 정확도 검증을 진행하여 객체인식 모듈 실행한다. 먼저, 객체 검출을 실행한다. 객체 검출은 영상속에서 개인의 민감 데이터인 얼굴, 차량 번호판을 인식한다. 다음으로 객체정보의 정확도를 표시한다. 원거리, 정면, 측면, 마스크 및 모자 착용 모두 인식한다. 검출된 객체들을 Batch De-Identification Tool 을 사용하여 개인정보 비식별화 처리기법 중 데이터 마스킹 기법으로 이미지 파일은 내용 전체에 암호화를 적용할 필요가 없기 때문에 임의잡음추가 기술을 이용하여 민감한 개인정보에 임의의 값을 곱하거나 더하여 개인 식별 정보의 노출을 방지한다.

얼굴 비식별화 예시는 '그림 18'은 원천 데이터, '그림 19'는 비식별화 조치된 영상이다

얼굴 비식별화 예시

Figure 18.Raw Data
그림 18.원천데이터

Figure 19.De-identification Example
그림 19.비식별화 예시

차량번호 비식별화 예시는 '그림 20'은 원천 데이터, '그림 21'는 비식별화 조치된 영상이다

차량번호판 비식별화 예시

Figure 20.RawData
그림 20.원천데이터

Figure 21.De-identification Example
그림 21.비식별화 예시

IV. 결론

본 연구에서는 소프트맥스와 ArcFace 손실함수를 활용한 k-익명성 중 데이터 마스킹 기법을 적용한 인공지능 학습 데이터의 개인정보 비식별화 자동화 도구를 개발하였다. 이를 통해 대규모 얼굴인식에서도 안정적으로 비식별화가 진행되도록 하였으며, 이를 다양한 분야의 인공지능 학습용 데이터 구축사업에서 활용할 수 있을 것으로 기대된다. 특히, 이미지나 동영상 원본을 획득한 후 본 모델을 적용하여 개인정보 비식별화를 수행하고 그 산출물을 라벨링하여 원천데이

터의 개인정보 재활용을 차단할 수 있을 것으로 판단된다. 또한, 이러한 기술을 다양한 온라인 콘텐츠에서 활용하여 거리 지도, SNS 동영상, 홍보 게시물 등에서 노출이 우려되는 개인정보를 삭제할 수 있는 API를 개발하고 이를 온라인 서비스로 제공함으로써 개인정보 침해 요소를 사전에 제거하는 서비스를 제공할 수 있을 것으로 기대된다. 향후 길거리 사진이나 문서에서 주소, 전화번호, 카드번호, 성별 등 개인정보 요소를 자동으로 탐지할 수 있는 모델을 개발하여 보다 효과적으로 개인정보 보호를 실현할 수 있는 방안을 연구하고자 한다.

V. 참고문헌

[1] S. D. Moon, "A study on the reformation of personal information protection law in the era of the 4th industrial revolution: focused on three data laws.", Domestic master's thesis Hanyang University Graduate School of Public Policy, 2021.
[2] J. H. Lee, “Crude oil of the 4th industrial revolution, ‘data specialist company’ to lead the revitalization of the data economy, on-site visit and conference held”, Ministry of Science and ICT, press release, Apr. 2018.
[3] S. J. Jong, “Legal review on protection and use of the personally non-identifiable information”, 2010.
[4] D. H. Kim, S. S. Kim, “A New Scheme for Risk Assessment Based on Data Context for De-Identification of Personal Information”, Journal of The Korea Institute of Information Security & Cryptology, VOL.30, NO.4, Aug. 2020.
[5] J. H. Jang, Y. J. Gim, “Policy direction for improving the effectiveness of AI learning data business”, NIA, IT & Future Strategy Report, 2020.
[6] W. J. Moon and 7 others, “Effect of Machine Learning Education Focused on Data Labeling on Computational Thinking of Elementary School Students”, Journal of The Korean Association of Information Education Vol. 25, No. 2, pp. 327-335, April 2021.
[7] H. C. Yang and 4 others, “A Guide to Using Personal Information De-identification Technology for Big Data Utilization ver 1.0”, NIA, 2015.
[8] S. T. Oh and 6 others, “Video data anonymization technology and evaluation method”, NIA, 2019
[9] H. W. Jung, “Support to strengthen data privacy protection for artificial intelligence (AI) learning”, Personal Information Protection Committee, press release, Jun, 2021.
[10] Mei Wang, Weihong Deng, Deep Face Recognition: A Survey, In Neurocomputing 14 March 2021 429:215-244
[11] Jiankang Deng , Jia Guo, Niannan Xue, Stefanos Zafeiriou,(2019) ArcFace: Additive Angular Margin Loss for Deep Face Recognition
[12] MAXTED Co., Ltd, "Web-based artificial intelligence data labeling service: Max Data Platform”, 2021
[13] Alejandro Peña and 4 others, “Facial Expressions as a Vulnerability in Face Recognition”, 2020

저자소개

이현주 (*Hyunju Lee*)

2021 년 8 월 동국대학교 대학원 기술창업학과 박사과정 수료
2021 년 9 월 ~ 현재 동국대학교 대학원 겸임교수
2022 년 12 월 ~ 현재 맥스테드 데이터사업본부 본부장

관심분야 : 인공지능, 비식별화, 생성형 인공지능, IT 융합

이승엽 (*Seungyeob Lee*)

2014 년 10 월 : 필리핀 EARIST 국립대학교 IT 학 석사
2020 년 2 월 : 동국대학교 일반대학원 기술창업학 박사수료
2010 년 8 월 ~ 현재 : 주식회사 맥스테드 대표이사

관심분야 : 인공지능, 모빌리티, 사물인터넷

전병훈 (*Byunghoon Jeon*)

2000 년 3 월- 2005 년 2 월 중부대학교 전기공학과 교수
2005 년 3 월- 현재 동국대학교 전자전기공학과 교수
2016 년 3 월- 현재 동국대학교 일반대학원 기술창업학과 교수

관심분야 : 기술사업화, IT 융합, 기술사업화