

일반논문 (Regular Paper)

방송공학회논문지 제27권 제5호, 2022년 9월 (JBE Vol.27, No.5, September 2022)

<https://doi.org/10.5909/JBE.2022.27.5.794>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

새로운 반려견 등록방식 도입을 위한 안면 인식 성능 개선 연구

이 동 수^{a)}, 박 구 만^{a)†}

A Study on Improving Facial Recognition Performance to Introduce a New Dog Registration Method

Dongsu Lee^{a)} and Gooman Park^{a)†}

요 약

동물보호법 개정에 따라 반려견 등록이 의무화 되었음에도 불구하고, 현재 등록 방법의 불편함으로 등록율이 저조한 상태이다. 본 논문에서는 새로운 등록 방법으로 검토되고 있는 반려견 안면 인식 기술에 대한 성능 개선 연구를 진행하였다. 딥러닝 학습을 통해, 반려견의 안면 인식을 위한 임베딩 벡터를 생성하여 반려견 개체별로 식별하기 위한 방법을 실험하였다. 딥러닝 학습을 위한 반려견 이미지 데이터셋을 구축하고, InceptionNet과 ResNet-50을 백본 네트워크로 사용하여 실험하였다. 삼중항 손실 방법으로 학습하였으며, 안면 검증과 안면 식별로 나뉘어 실험하였다. ResNet-50 기반의 모델에서 최고 93.46%의 안면 검증 성능을 얻을 수 있었으며, 안면 식별 시험에서는 rank-5에서 91.44%의 최고 성능을 각각 얻을 수 있었다. 본 논문에서 제시한 실험 방법과 결과는 반려견의 등록 여부 확인, 반려견 출입시설에서의 개체 확인 등 다양한 분야로 활용이 가능하다.

Abstract

Although registration of dogs is mandatory according to the revision of the Animal Protection Act, the registration rate is low due to the inconvenience of the current registration method. In this paper, a performance improvement study was conducted on the dog face recognition technology, which is being reviewed as a new registration method. Through deep learning learning, an embedding vector for facial recognition of a dog was created and a method for identifying each dog individual was experimented. We built a dog image dataset for deep learning learning and experimented with InceptionNet and ResNet-50 as backbone networks. It was learned by the triplet loss method, and the experiments were divided into face verification and face recognition. In the ResNet-50-based model, it was possible to obtain the best facial verification performance of 93.46%, and in the face recognition test, the highest performance of 91.44% was obtained in rank-5, respectively. The experimental methods and results presented in this paper can be used in various fields, such as checking whether a dog is registered or not, and checking an object at a dog access facility.

Keyword : Deep Learning, Face Recognition, Dog Recognition, ResNet, Triplet Loss

Copyright © 2022 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

2021년부터 발효된 동물보호법 개정에 따라 현재 국내에서 양육되는 반려견은 모두 농림축산식품부의 반려동물보호시스템에 개체 및 소유주 등록을 의무화하고 있다. 이 경우 동물 등록과 함께 등록된 반려견에 대한 개체 식별을 위해, 내장형 칩 또는 외장형 칩 등 두 가지 방법을 사용하고 있다. 내장형 칩의 경우, 동물병원에서 시술이 이루어지며, 분실이나 파손의 위험이 없는 반면 반려견의 피부에 이식하는 방식을 소유주들이 기피하는 경향이 많고, 상대적으로 비용이 비싸다. 외장형 칩의 경우, 동물 등록 대행업체를 통해 주로 목걸이 형태로 판매되고 있어 편리한 점이 있으나, 분실이나 파손이 발생하기 때문에 유실견이 발생했을 때 개체를 확인할 수 없는 단점이 있다.

이에 대한 대안으로 농림축산식품부에서는 수년전부터 인공지능의 인식 기술을 이용한 반려동물 등록 방식을 고려해오고 있으나, 반려동물의 개체 식별에 대한 인식 성능이 보장되지 못하고 등록하는 영상의 품질이 일정하지 않아 공식적인 방법으로 도입하지 못하고 있다. 딥러닝을 이용한 영상 인식 기술이 비약적인 발전을 이루면서 사람에 대한 자동 판별 성능은 인간의 식별 수준을 능가하였으나, 동물에 대한 인식 기술은 상대적으로 매우 미흡한 상태이다.

반려견으로 범위를 좁혀서 보면, 인간에 대한 얼굴 인식 문제와 달리, 반려견의 안면 인식 문제에서 가장 큰 어려움은 활용 가능한 데이터셋 확보가 어렵다는 점이다. 반려견의 안면 인식 응용 분야중에서 반려견의 자동 품종 분류를 위한 인식 문제의 경우, 웹 또는 기존 연구 데이터로부터 학습 데이터셋을 일정 분량 이상 수집하는 것이 가능하다. 또한 대량의 데이터를 기반으로 결과가 검증된 우수한 분

류 모델을 활용하여 전이학습(Transfer Learning)을 수행함으로써, 상용화 수준에 근접한 인식 성능을 얻을 수 있다^{[1][2][3]}. 반면, 반려견의 개체별 식별을 위해서는 각 개체별로 다수의 이미지가 대량으로 확보되어야 하므로 학습 데이터를 확보하는 것이 쉽지 않다. 또한, 사람의 얼굴과 달리 형상이 복잡하고 변수가 많아 개체 식별을 위한 인식 성능을 높이는 데 어려움이 있다.

본 논문에서는 딥러닝 학습을 통해 반려견의 개체 식별 모델을 만들고 반려견에 대한 안면 인식 성능을 개선하기 위한 실험적 방법을 제안하였다. 반려견 안면 인식의 제한 사항을 극복하기 위해서는 소량의 데이터를 최대한 활용할 수 있는 적절한 전이학습의 개발, 최적의 하이퍼 파라미터 탐색, 인식 알고리즘과 손실함수의 선택 등 핵심적인 사항을 고려해야 한다.

이를 위해서 본 연구에서는 반려견 안면 인식 모델을 통해 개체별 특징을 나타내는 임베딩 벡터 값들을 구축하고, 핵심 파라미터의 변화에 따른 실험 결과를 통해 우수한 성능을 가지는 인식 모델을 확보함으로써, 반려견의 안면 인식을 이용한 등록 방법의 기반을 만들고자 한다. 이 연구의 결과물은 반려동물보호시스템으로의 등록을 위한 사용자 인터페이스 웹 또는 앱과 연동되어, 새로운 반려견 등록 방식으로 사용이 가능할 것으로 기대된다.

본 실험에서는 사람의 얼굴 인식에서 획기적 성능을 보여 주면서 가장 우수한 손실함수로 인정받는 구글의 삼중항 손실(Triplet Loss) 방식에 대해 반려견의 이미지를 적용하여 인식 성능을 실험하였다. 이와 유사한 실험은 여러 해외 연구에서 시도하였으나^{[4][5]}, 동일 반려견에 대한 학습 데이터셋의 제약과 하이퍼 파라미터의 조합이 다양하기 때문에 결과치의 범위가 넓게 나타나고 있으며, 아직 개선의 여지가 있는 것으로 생각된다. 동일한 데이터셋에 대해서 임베딩 벡터값을 산출하는 네트워크 모델의 선택도 중요하며, 네트워크의 하이퍼 파라미터 외에 삼중항 학습을 위한 데이터셋을 선택할 때 hard triplet 비중, 삼중항 마진(Triplet Margin) 변화 등에 따라서 성능이 다르게 나타나는 걸 볼 수 있다.

본 실험에 사용한 반려견 이미지는 2019년 G. Mougeot가 반려견 인식 연구를 위해 사용했던 1,393마리의 사진

a) 서울과학기술대학교 나노IT디자인융합대학원(Graduate School of Nano IT Design Fusion, Seoul National University of Science and Technology)

‡ Corresponding Author : 박구만(Gooman Park)

E-mail: gmpark@seoultech.ac.kr

Tel: +82-2-970-6430

ORCID:https://orcid.org/0000-0002-7055-5568

· Manuscript August 16, 2022; Revised September 1, 2022; Accepted September 1, 2022.

8,360장으로 기본 데이터셋을 만들고^[6], 컬러 지터(color jitter), 수평 플립, 랜덤 로테이션(random rotation) 등 데이터 증강 기술을 사용하여 실험 데이터셋을 구축하였다. 일반적으로 개체 식별을 위한 안면 인식의 평가에서는 안면 검증(verification)과 안면 식별(recognition)으로 나뉘어 수행하게 되며, 두 가지 기술이 각각 응용서비스로 활용될 수 있다. 전자는 1:1로 비교하는 안면 검증이며, 후자는 1:N으로 DB에서 검색하는 안면 식별을 의미한다. 본 연구의 1차 실험에서는 FaceNet에서 사용했던 InceptionNet을 백본 네트워크로 하고, 삼중항 데이터를 구성하여, 삼중항 손실로 학습하였다. 파라미터의 변화에 따라 안면 검증에서는 정확도 86.22~90.22%의 정확도를 달성하였고, 안면 식별에서는 rank-1 정확도로 71.51~73.6%를 기록하였다.

2차 실험에서는 Inception 모델 대신 ResNet-50 모델을 백본 네트워크로 사용하여, 사전학습된 가중치를 사용한 경우와 새롭게 가중치를 생성한 경우에 대해 각각 파라미터를 변경하면서 학습을 수행하였다. 이 실험에서 주요 파라미터 변화로 삼중항 마진 값, hard triplet 비율, 안면 검증을 위한 임계값(threshold) 값을 조정한 결과, 안면 검증에서는 최고 93.46%의 정확도를 얻었으며, 안면 식별에서는 rank-1 최고 81.15%를 얻음으로써, 성능이 크게 향상됨을 확인하였다.

반려견의 안면 식별 응용을 고려할 때, 반려견 등록을 위해 본 모델을 사용하여 인식용 임베딩 DB를 구축한 후, 유기견에 대한 조기 확인, 반려견 시설의 출입 인식기, 반려견 등록 여부 등으로 활용될 예정이므로, 1:N 방식의 인식에서는 rank-5 결과값도 의미있는 활용 값이 된다. 안면 식별에서 rank-5 기준으로는 2차 실험에서 최고 91.44%의 성능을 확인하였다. 이 결과는 반려견의 안면 검증 성능으로는 국내외 선행 연구와 비교하여 최고 수준이라 할 수 있으며, 향후의 활용에 따라 충분히 상용화 적용이 가능한 성능이라 할 수 있다.

본 논문 2장에서는 딥러닝 기반의 얼굴인식 기술에 대한 핵심요소와 본 연구와 밀접한 관련이 있는 연구 결과를 검토하였고, 특히 반려동물에 대해 적용한 기술을 살펴 보았다. 3장에서는 반려견에 대한 삼중항 데이터셋 구축 방법, 4장에서는 안면 인식을 위한 실험 내용과 결과를

기술하였으며, 5장에서 결론과 향후 활용 계획에 대해 기술하였다.

II. 관련 연구

1. 관련 연구동향

딥러닝에 의한 데이터 학습 기반으로 디지털 이미지를 처리하기 이전에는, 일반적으로 feature engineering 방식으로 알려져 있는 SIFT(Scale-invariant feature transform)^[7], SURF(Speeded up robust features)^[8] 등과 같은 기술을 사용하여 이미지로부터 특징점을 추출하는 방법들이 사용되었다. 이러한 방법들은 입력 이미지의 차원을 줄이고 특징 정보를 추출한 다음에, 효과적인 분류 또는 회귀 결과를 얻기 위해 SVM(Support Vector Machine) 등의 기술을 다시 적용해야 한다. 따라서 이미지 처리를 통해 결과를 도출하기까지 프로세스가 복잡한 과정을 가지게 되지만 딥러닝 기술이 증명되기 전까지는 이러한 feature engineering 방법이 가장 앞선 기술들이라 할 수 있다.

2012년 ImageNet 챌린지에서 AlexNet^[9]이 CNN(Convolutional Neural Networks) 기반의 딥러닝 방식으로 압도적인 성능을 얻은 후로 딥러닝 영상 처리 분야에서는 딥러닝에 의한 데이터 학습 방법이 전통적인 feature engineering 방식을 대체하게 되었다. 이후 딥러닝 방식은 네트워크의 깊이를 증가하고 구조를 다양화하는 동시에, 데이터를 보다 효과적으로 학습시킬 수 있는 손실함수(loss function)들이 개발되어 오차를 줄일 수 있었다. 최근에는, 시계열 모델에 사용되었던 트랜스포머(transformer) 기술을 영상처리에 적용한 비전 트랜스포머 모델이 다양하게 연구되고 있으나, 대규모 학습 데이터와 대용량 컴퓨팅 파워를 기반으로 하는 점에서 제한사항이 있다.

딥러닝 기술을 이용한 얼굴 인식 분야에서는 2014년 Y. Taigman 등이 발표한 논문^[10]에서 인간의 얼굴 인식 분야에서 획기적인 전기를 갖게 된 이후, 새로운 모델과 손실함수의 개발을 통해 성능이 지속적으로 개선되었으며, 인간의 얼굴에 대한 대규모의 학습 데이터를 기반으로 이미 인간보다 뛰어난 인식 성능을 가진 기술이 개발되어 상용

화되고 있다.

하지만, 인간의 얼굴 인식과 달리 동물의 안면 인식은 형태의 복잡성으로 인해 매우 어려운 작업이다. 특히 반려견의 경우 개체간 차이가 매우 미세하고, 학습을 위해 필요한 대규모의 적절한 데이터셋 확보도 어렵기 때문에 개체간 식별에 대한 연구는 많지 않으며, 성능 측정도 제한 조건을 가진 환경에서 이루어지고 있는 상황이다.

2. 딥러닝 기반 얼굴 인식

인간의 얼굴 검증(verification) 및 인식(recognition)에 대한 연구는 오랫동안 반대하게 수행되어 온 분야이며, 특히 딥러닝 기술을 사용하면서 전통적인 패턴인식 방법 대비 성능이 획기적으로 개선되었다.

Y. Taigman 등은 2014년 CNN(Convolutional Neural Network) 모델을 변형한 9 계층의 딥러닝 모델을 적용하여 기존의 특징 기반(feature engineering) 얼굴 인식 방식에서 가졌던 오류를 27%이상 줄임으로써, 인간의 안면 검증에서 획기적인 발전을 보여주었다^[10]. DeepFace 기술로 명명한 이 시스템에서는 페이스북에서 수집된 440만장의 대규모 얼굴 이미지를 사용하였다(SFC: Social Face Classification 데이터셋). 이 논문에서는 딥러닝 학습을 위해 얼굴에 67개의 랜드마크를 설정하고 3D 정렬 기술로 정면화(frontalization) 전처리를 진행한 후, 딥러닝 네트워크에 넣어 얼굴 인식을 수행하였다.

이 논문에서는 LFW(Labeled Faces in the Wild) 데이터

셋 기준 얼굴 비교검증에서 97.35%의 얼굴 인식 정확도를 달성하였고, YTF(YouTube Faces) 데이터셋에서도 91.4%의 정확도 성능을 보여줌으로써, 컴퓨터 비전 분야에서 얼굴 인식 연구를 딥러닝 기반으로 전환하는 계기를 만들었다.

2015년에 구글의 F. Schroff 등은 FaceNet 시스템 논문에서 얼굴 인식 학습을 위한 새로운 삼중항 손실(Triplet Loss) 방법을 개발하였다^[11]. 삼중항 손실은 그림 2에 표시한 것처럼 하나의 기준 이미지(Anchor)를 두고 동일인의 다른 사진(Positive)과 다른 사람의 사진(Negative)을 하나의 데이터셋 단위로 설정한 다음, 각 이미지로부터 추출된 임베딩 벡터값을 사용하여 anchor와 positive 간은 벡터간 유클리드 거리가 최대한 가깝도록 하고 anchor와 negative 간은 거리가 멀어지도록 학습시키는 방법이다. 이를 위한 손실 함수는 식 1과 같으며, 이를 최소화시키는 방식으로 학습된다.

여기서, $f(x_i^a)$, $f(x_i^p)$, $f(x_i^n)$ 는 각각 anchor 이미지, positive 이미지, negative 이미지에 대한 임베딩 벡터 값을 나타내며, α 값은 서로 다른 클래스 간 차이를 만들기 위한 마진(margin) 을 나타낸다.

이 논문에서는 최대 약 8백만명의 사람들에 대한 사진 2억 여장을 학습 데이터셋으로 사용하였으며, 학습 데이터셋에 대해 다양한 크기의 크기변환 이외의 별도로 전처리를 사용하지 않았다. 백본 네트워크로는 Zeiler&Fergus Net^[12] 및 Inception Net^[13] 기반 CNN 네트워크 4종류를 사용하여 실험하였다. CNN 네트워크에서는 각 이미지에 대한 128차원

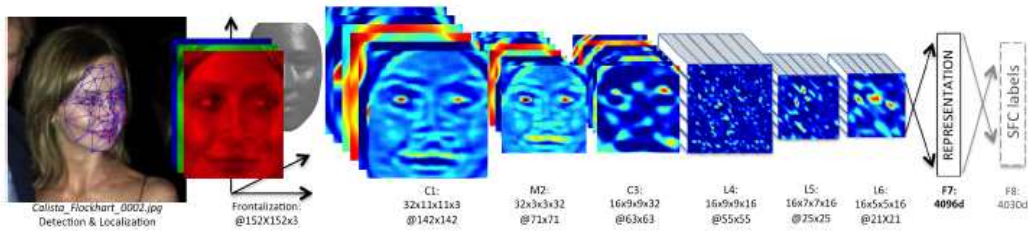


그림 1. DeepFace 네트워크 아키텍처 요약^[10]
 Fig. 1. Outline of the DeepFace Network Architecture^[10]

$$L = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|^2 - \|f(x_i^a) - f(x_i^n)\|^2 + \alpha] \quad (1)$$

임베딩 벡터를 추출한 후, 삼중항 손실로 학습하는 방법을 수행하였다. 벤치마크 데이터셋인 LFW에서 실험한 결과, 당시 SOTA(State-of-the-Art)였던 DeepID2+[14] 보다 오류율을 30%까지 줄이면서 1:1 안면 검증에서 99.63%의 놀라운 성능을 달성하였고, YTF에서는 95.12%의 정확도 성능을 보여주었다.

이후 얼굴 인식 분야의 성능 개선 방법으로는 학습을 정교하게 하기 위해 새로운 손실함수를 개발하는 연구가 이루어졌다. 유클리드 거리 기반의 메트릭 학습 대신 Angular Softmax[15]나 Cosface 방법[16] 등이 개발되었다. Cosface 논문에서 저자들은 서로 다른 클래스 간에 cosine margin 값을 더 크게 만드는 Large Margin Cosine Loss(LMCL) 방법을 개발하였다. 벤치마크 데이터셋으로 실험한 결과, LFW 데이터셋으로 99.73%의 정확도 성과와 YTF 데이터셋에서 97.6%의 정확도 성능을 각각 달성함으로써, 인간의 얼굴에 대한 인식으로는 사람을 능가하는 수준으로 발전하였다.

3. 동물에 대한 생체 인식

사람의 얼굴 인식에 대한 연구 성과와는 달리, 국내외에서 반려동물에 대한 생체 인식을 통해 개체를 식별하는 연

구는 매우 적은 편이다.

딤러닝 방식이 적용되기 이전에는 영상처리 기술을 사용하여 주로 동물의 개체마다 고유한 코 부분(muzzle)의 패턴을 사용하는 연구가 시도되어 왔다. 대표적으로 국소 특징점 추출방식의 하나인 SURF(Speeded Up Robust Feature)와 RANSAC(Random Sampling Consensus) 알고리즘을 활용하여, 코 부분의 패턴을 비교함으로써 개체를 인식하는 방법이 있다[17]. 이 방법은 소수의 개체 검증(verification) 또는 식별(recognition)에서는 비교적 높은 성능을 얻을 수 있는 방법이나, 고해상도의 근접 이미지 포착이 어렵고 일상적인 필드 환경에 있는 동물 이미지에 대해서는 적용하기 어려워 범용적인 방법으로 사용하기에는 적합하지 않다.

해외에서는 초기연구로서, 콜롬비아 대학의 J. Liu 등이 개의 품종별로 얼굴 모형의 기하학적 모델을 구축하여 품종별 분류를 시도했다[18]. 그들은 Flickr, Google 사이트 등에서 총 133 품종의 개 사진 8,351장을 수집하여 데이터셋을 구성한 후, 그림 3에 보는 것처럼 SIFT(Scale-Invariant Feature Transform) descriptor를 사용하여 8개의 포인트로 외형 특징을 추출하였다. 분류에는 SVM(Support Vector Machine) 기반의 검출기를 설계하였으며, 67%의 품종 분류 정확도를 보여주었다.

딤러닝을 이용한 인식 작업의 경우, 국소 특징점을 추출

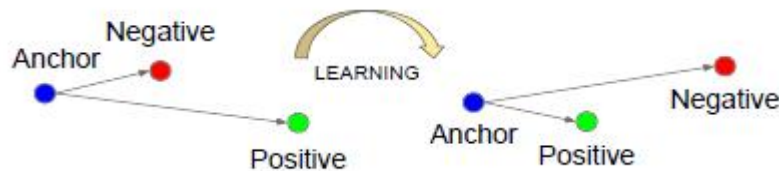


그림 2. 삼중항 손실 학습[11]
Fig. 2. The Triplet Loss Learning[11]



그림 3. 8개 부분으로 라벨된 이미지 샘플[18]
Fig. 3. Image samples labeled with 8 parts[18]

하기 위한 영상처리과정 보다는 기본적으로 충분한 데이터셋의 확보가 전제되어야 하고 적절한 네트워크 모델의 설계와 학습을 위한 손실함수의 설정이 중요한 요소가 된다. 딥러닝을 적용한 개의 안면 인식 연구중에서 G. Mougeot 등은 FaceNet의 삼중항 손실을 사용하여 ResNet-like 모델과 VGG-like 모델에 적용하는 방법을 발표했다^[4]. 이들은 485마리의 개 이미지 3,148 장의 비교적 적은 데이터셋을 사용하였는데, 이를 보완하기 위해 얼굴의 세 포인트를 기준으로 얼굴의 각도를 변화하는 전처리를 수행하고, 삼중항 데이터셋을 자신들만의 방식으로 재배열하여 실험하였다.

그림 4에 표시한 ResNet-like 모델을 사용하였을 때, 안면 검증에서는 최고 92%의 정확도를 달성하였고, 안면 식별에서는 Top-1에서 60.4%, Top-5에서 88.4%의 성능을 보여주었다. 안면 식별을 위해서는 개체별 추출된 임베딩 벡터값을 기준으로 kNN(k Nearest Neighbor) 분류 알고리즘을 사용하였다. 딥러닝 기술을 개의 안면 인식 응용에 적용했을 경우의 가능성과 어려움을 보여준 실험 결과로 상당히 의미있는 연구라 할 수 있으며, 논문 발표 후에 데이터셋을 추가하여 오픈^[6]하기도 했다.

위와 유사한 실험으로 스탠포드 대학의 A. Klein은 딥러닝의 객체 검출 기술과 EfficientNet 등 몇가지 네트워크 모

델을 이용하여 고양이 얼굴에 대한 인식 연구를 수행하였다^[19]. 개체당 5장 이상의 사진을 가지는 고양이 사진 약 96,000장의 사진을 데이터셋으로 사용하였으며, 이전의 연구와 달리 랜드마크 태깅이나 이미지 정렬 등의 전처리를 사용하지 않고, YOLOv3 객체 검출 알고리즘을 사용하여 고양이 얼굴에 대한 검출을 위해 사전학습을 수행하였다. 이 과정에서는 적절한 confidence threshold를 찾아서 mAP(Mean Average Precision) 성능을 높이는 것이 중요하다. 최종 안면 식별을 위한 학습방법으로는 삼중항 손실을 사용하였으며, 백본 네트워크로는 EfficientNet-B2를 적용하여 실험하였다. 이 실험에서는 안면 검증에서 최적의 L2 거리 임계값을 0.455로 찾아서 최고값 95%의 정확도로 SOTA 성능을 달성하였다. 하지만 표 1에서 보는 것처럼, 개체 식별을 위한 안면 인식에서는 Top-1에서 47.45%,

표 1. 고양이 얼굴 인식 실험 결과
 Table 1. Accuracy results of cat face identification

M	k	Identification Accuracy		
		Mean	Min	Max
1	1	47.45%	45.18%	49.42%
2	3	70.78%	68.97%	72.52%
3	4	77.19%	75.15%	79.28%
4	5	81.32%	78.89%	83.95%

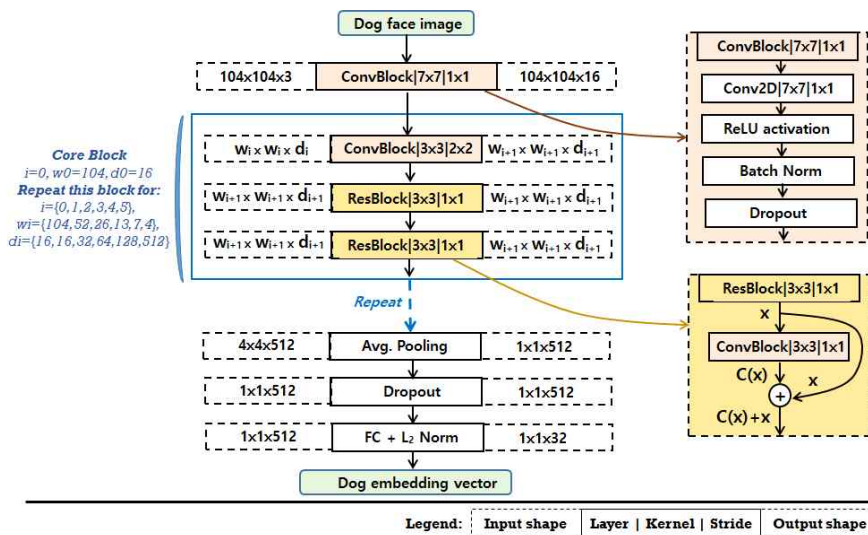


그림 4. ResNet-like 모델 정의^[4]
 Fig. 4. ResNet-like model definition^[4]

Top-5에서 81.32% 성능을 나타냄으로써, 여전히 인식의 어려움이 남아있음을 알 수 있다. 저자는 딥러닝 학습 방법론보다 우수한 고품질의 입력 데이터 DB가 있다면 더 좋은 성능을 가질 수 있다고 설명한다.

III. 삼중항 데이터셋 구축

삼중항 손실을 사용하여 반려견 안면 인식을 위한 학습용 데이터셋을 구축하기 위해서는, 최적의 학습 효과를 얻을 수 있도록 일반적인 삼중항 데이터와 Hard triplet 데이터를 적절한 비율로 구성하여 입력 데이터를 새롭게 구축해야 한다. 먼저 기본적인 반려견 학습 데이터셋을 얻기 위해 2장에서 기술한 관련 논문과 Kaggle 사이트로부터 반려견 이미지 데이터를 분석한 결과, G. Mougeot 등이 연구한 논문에서 사용한 데이터셋이 가장 좋은 품질의 반려견 이미지를 제공한다고 판단하였다. 특히 총 8,360장의 데이터는 각 개체별로 2~14장의 jpeg 이미지를 가지는데, 각 이미지는 배경의 노이즈를 최소화하고 얼굴 부분을 중심으로 다양한 각도와 포즈를 포함하고 있어, 학습 데이터로 사용하기에 적합하다. 그림 5에 나타난 것처럼 학습 데이터는 총 1,393마리에 대해 일련번호로 구성되며, 폴더 이름이 라벨이 되도록 구성되어 있다.

본 연구의 실험을 위해서는 입력 데이터를 삼중항 데이

터로 만들어야 한다. 즉, 한 장의 앵커(Anchor) 이미지와 동일 개체의 다른 사진(Positive), 다른 개체의 사진(Negative)로 이루어진 세 장의 이미지가 한개의 데이터셋이 된다. 각각의 입력 단계에서 랜덤하게 선택해서 배치(Batch)로 구성하고 네트워크를 통과하게 되면, 임베딩 벡터 $f(x_a)$, $f(x_p)$, $f(x_n)$ 이 각각 만들어진다. 삼중항 손실에 의한 학습은 동일한 개체에 대한 벡터간 유클리드 거리($\|f(x_a)-f(x_p)\|$)는 더 가깝게 만들고, 서로 다른 개체의 벡터간 거리($\|f(x_a)-f(x_n)\|$)는 더 멀어지게 되도록 만드는 것이 목표이다.

하지만 그림 6에서 보는 것처럼, 딥러닝 모델을 거쳐서 나온 임베딩 벡터의 경우, 대부분 동일한 개체의 벡터간 거리는 가깝고, 서로 다른 개체의 벡터간 거리는 멀게 되므로 삼중항 손실 식을 쉽게 만족하게 된다. 이러한 임베딩 벡터 결과값을 가지는 삼중항 데이터로는 마진 값을 두더라도 좋은 학습 효과를 기대하기 어렵다.

이를 보완하기 위한 방법은 Hard triplet을 구성하는 것이다. Hard triplet은 삼중항 데이터를 구성할 때, 동일한 개체간 벡터는 최대한 먼 이미지(hard positive)를 선택하고, 서로 다른 개체간 벡터는 최대한 가까운 이미지(hard negative)를 선택함으로써, 모델로 하여금 학습을 어렵게 만드는 방법이다. 즉 그림 7에서 보면 앵커 이미지를 중심으로 hard positive와 hard negative 이미지를 선택해서 Hard triplet 데이터셋을 구성하여 입력데이터로 사용하게 된다.

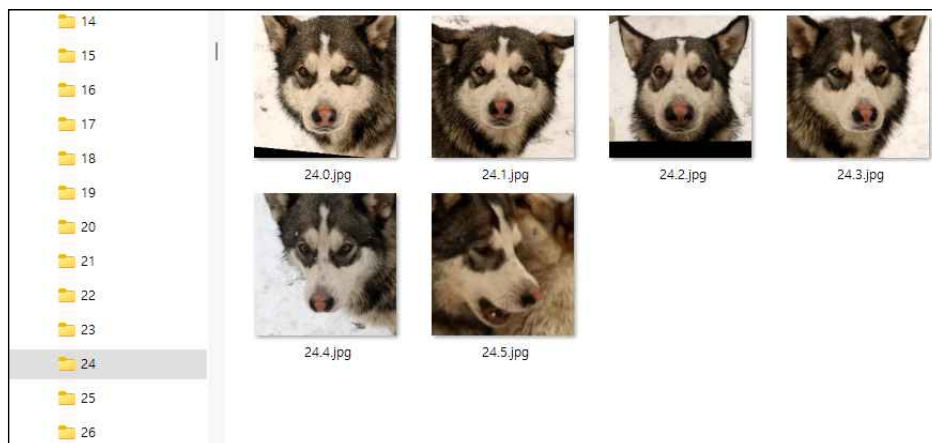


그림 5. 폴더 이름으로 라벨링된 데이터셋 샘플
Fig. 5. Dataset sample labeled by folder name

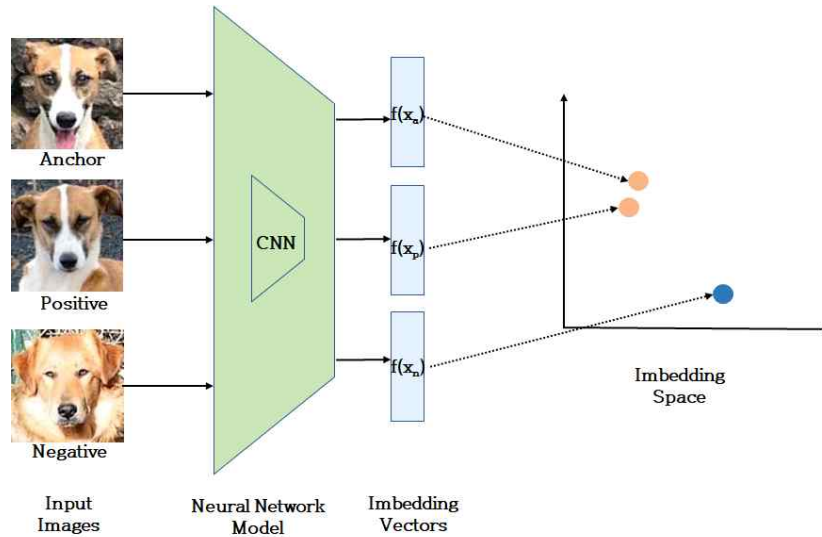


그림 6. Anchor, positive, negative 이미지에 대한 임베딩 벡터 결과
 Fig. 6. Imbedding vector values for the images of anchor, positive and negative

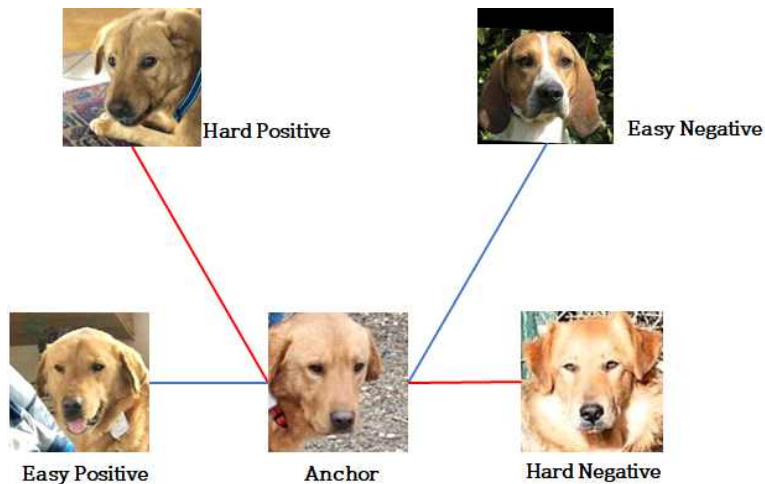


그림 7. 하드 퍼지티브와 하드 네가티브 이미지 샘플
 Fig. 7. Image samples of hard positive and hard negative

하지만 모든 입력데이터를 Hard triplet으로만 구성하면 딥러닝 모델은 학습을 진행하지 못하게 된다. 이를 해결하기 위해서는 모델이 학습을 효과적으로 할수 있도록 입력 데이터에 적절한 비율로 Hard triplet을 구성해야 한다. 이는 휴리스틱한 방법으로 값을 찾아야 하며, 본 실험에서는 이 비율을 20%로 했을때 최상의 결과를 얻을수 있었다.

IV. 안면 인식 실험

1. 안면 인식 네트워크 모델의 선택

본 연구에서 안면 인식 네트워크 모델은 관련된 선행 연구에서 가장 많이 사용한 InceptionNet과 ResNet에 대해 각 네트워크의 특성에 맞추어 파라미터를 변경하면서 실험을

진행하였다.

InceptionNet은 2014년 ImageNet 이미지 분류 챌린지에서 우수한 모델이기도 하지만, 구글 엔지니어들이 삼중합 손실을 개발하여 FaceNet 얼굴인식 모델에도 적용했던 모델이다. InceptionNet의 특징은 CNN의 깊이를 더욱 깊게 하기 위해 별도의 인셉션 모듈을 만들어서 연산량을 매우 효율적으로 감소시킨 것이 특징이다. 인셉션 모듈은 자체로 컨볼루션 연산을 하지만 채널의 수를 줄이는 효과를 가질 수 있으므로, 그림 8에서처럼 InceptionNet에서는 총 9개의 인셉션 모듈을 사용하였고, 전체적으로는 풀링 계층을 포함하여 총 27계층의 매우 깊은 신경망 구조를 가지면서도 기울기 소실 문제를 해결할 수 있었다.

ResNet의 네트워크 구조는 VGGNet 모델을 기반으로 하여, 컨볼루션 계층을 대폭 증가시켜 더욱 깊은 신경망 구조를 가진다. 계층을 깊게 만들기 위해 저자들은 잔차 블록

(residual block)을 고안하여 학습시킨다. 그림 9에 나타난 것처럼 기존의 순차적인 입출력 구조와 달리, 네트워크의 입력 값을 복제하여 일정 크기의 컨볼루션 계층이 지날때마다 출력 값에 더해주는 skip connection 구조를 가진다. 본 실험에 적용한 ResNet-50에서는 3개의 컨볼루션 계층 단위로 skip connection이 일어난다.

그림 9에서 일반적인 컨볼루션 신경망의 출력은 $H(x)$ 가 되겠지만, ResNet 구조에서는 $H(x)=F(x)+x$ 가 된다. Identity mapping을 이용하는 구조인데, 입력값 x 가 출력값에 그대로 나오기 위해서는 $F(x)$ 를 0에 가깝게 만드는 것이 목적이 된다. $F(x)=H(x)-x$ 이므로 $F(x)$ 를 최소로 해준다는 것은 $H(x)-x$ 를 최소로 해주는 것과 동일한 의미를 가지며, 이 $H(x)-x$ 를 잔차(residual)라고 한다. 즉, 잔차 블록에서는 잔차를 최소화하도록 학습하게 된다.

ResNet의 여러 시리즈 중에서 ResNet-50 부터는 더 많은

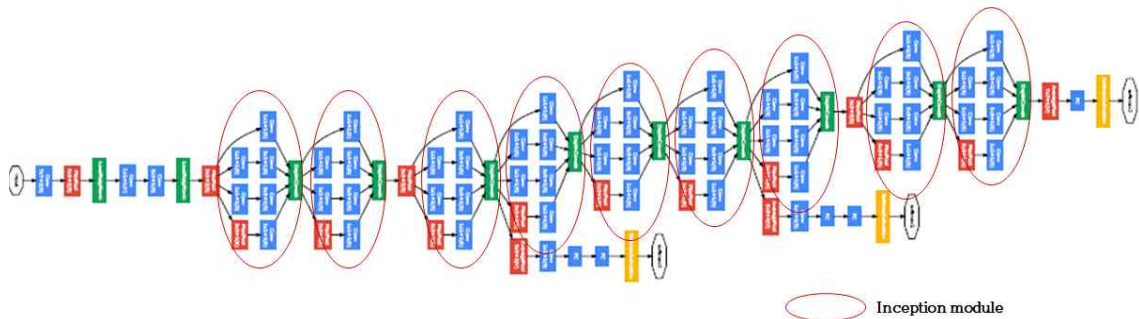


그림 8. 인셉션 모듈을 포함한 구글넷 구조
Fig. 8. GoogLeNet network architecture containing inception modules

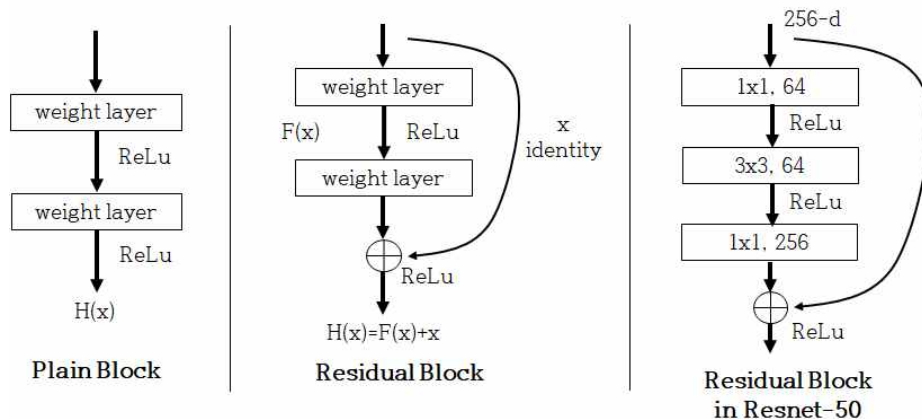


그림 9. Skip Connection을 가지는 잔차블럭
Fig. 9. Residual Block with Skip Connection

계층을 쌓기 위해서 그림 9에 표시한 것처럼 컨볼루션 계층이 3개가 있는 bottleneck 구조를 가진다. 계층이 깊어질수록 bottleneck에서는 256개, 512개, 1,024개, 2,048개의 출력 채널이 생성하게 되고, 마지막으로 average pooling을 거친 후 FC(Fully Connected) 계층으로 연결된다.

2. 실험 진행 및 결과

FaceNet에서 수행한 인간의 안면 인식 실험과 달리 반려견의 인식을 위한 이미지 데이터량은 매우 빈약하다. 이를 보완하기 위해 본 실험에서는 표 2와 같이 크게 3가지의 데이터 증강 기법을 적용하여 전처리를 수행하였으며, 원본 이미지의 품질을 고려해서 color jitter에서는 작은 범위의 값으로 조정하였다.

표 2. 실험에 적용된 데이터 증강 기법
 Table 2. Data augmentation methods applied on experiments

Augmentation	Values
Color jitter	bright=0.3, contrast=0.3, saturation=0.3, hue=0
Horizontal flip	0.3
Random rotation	-10~10도

본 연구에서는 이미지 분류를 위한 학습과 달리 안면 인식 성능에 영향을 미치는 몇가지 중요한 하이퍼 파라미터

를 설정하기 위해 다양한 실험을 진행하였다. 먼저, 선택한 백본 네트워크 모델이 효율적인 학습을 하기 위해서 중요하게 고려되어야 할 점은 삼중항 데이터셋에서 Hard triplet으로 구성되는 데이터셋의 비율을 적절히 설정하는 것이다. 삼중항 손실 연산에 사용되는 마진값도 학습에 영향을 미친다. 또한 1:1 안면 검증에서는 임베딩 벡터간 거리의 임계값(Threshold) 설정에 따라 성능에 영향을 미치게 된다.

나머지 학습율(Learning Rate)이나 드롭아웃(Dropout) 비율 등 네트워크 모델과 관련된 하이퍼 파라미터들은 인식 성능에 직접적인 영향이 약하므로 기존 연구에서와 유사한 값으로 설정하였다. 학습시 배치 크기는 16으로 하였으며, Adam 옵티마이저를 사용하였고, 학습 에포크는 몇차례 실험 후에 손실 값이 안정화되는 200번으로 정하여 실험을 진행하였다.

2.1 InceptionNet 기반 실험

표 3은 InceptionNet을 사용하여 실험하였을 경우, 주요 하이퍼 파라미터의 설정 값에 따라 학습한 인식 성능 결과를 보여 준다.

Hard triplet 비율은 표 3에 나타난 값 외에 다른 값으로도 실험하였지만 좋은 성능은 얻을 수가 없었으며, Hard triplet 비율을 0.6 이상으로 설정했을 경우에는 모델이 학습을 진행하지 못하였다. 삼중항 마진은 FaceNet 논문에서와 동일하게 0.2에서 좋은 성능을 얻을 수 있었다.

서론에서 언급한 것처럼 본 연구 결과의 주요 활용 분야

표 3. InceptionNet 기반의 실험 결과 정리
 Table 3. Summary of InceptionNet based experiments

Items		Exp. 1	Exp. 2	Exp. 3	Exp. 4
Common values	Number of epochs	200			
	Dataset split	Train:Test = 7:3			
	Learning rate	1.00E-04			
	Dropout rate	0.5			
Hyper-parameters	Hard triplet ratio	0.4	0.2	0.2	0.2
	Triplet margin	1.0	1.0	0.5	0.2
	Threshold for verification	1.2	1.2	1.1	0.8
Accuracy results (%)	Verification	86.22	89.35	90.22	90.01
	Recognition_rank1	71.51	71.33	71.60	73.60
	Recognition_rank5	84.60	85.46	84.83	86.10

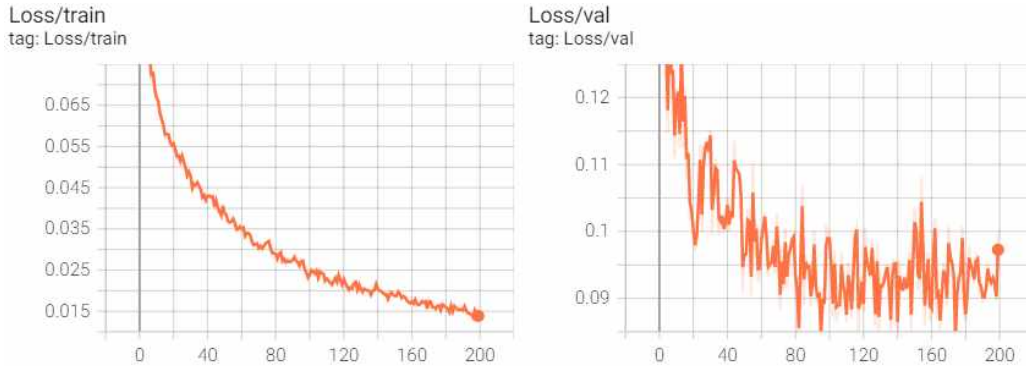


그림 10. 실험 4에 대한 손실값의 변화
Fig. 10. Loss values variation for Exp. 4

인 유실건의 조기 확인이나 출입시설의 인식기 응용 등을 고려할 때, 안면 식별 성능에서는 rank-5의 결과값도 유의미하게 활용될 수 있다. 그림 10은 표 3에서 가장 좋은 성능을 얻은 실험 4에 대한 학습중 손실값의 변화를 보여준다.

3. ResNet-50 기반 실험

ResNet-50을 백본 네트워크로 사용한 실험에서는 먼저 ResNet에서 사전학습된 가중치를 사용하지 않고 처음부터 새로운 가중치를 생성하면서 실험을 진행해본 결과 (실험 5), 사전학습된 가중치를 새롭게 업데이트하는 실험보다 상대적으로 성능이 좋지 않음을 확인할 수 있었다. 즉, 반려견 이미지 데이터로 새롭게 학습하더라도 ResNet에서 사전학습된 가중치가 특징 추출에 대한 능력을 우수하게 갖추고

있으므로 그 기반으로 가중치를 업데이트하는 실험 방법이 더 우수함을 확인할 수 있다. 표 4는 ResNet-50을 사용하여 실험하였을 경우, 주요 하이퍼 파라미터의 설정 값에 따라 학습한 실험 결과를 나타낸다.

표 4에서 볼수 있듯이 전체적으로 InceptionNet을 기반으로 한 실험에서보다 성능이 크게 향상되었음을 알 수 있다. 안면 식별 실험에서 rank-5의 최고 성능 91.44%를 얻음으로써, 추가적인 데이터셋 수집과 학습을 통해 반려견 등록을 위한 임베딩 벡터 구성에 대한 성공 가능성을 보여준다고 할 수 있다. 특히 1:1 안면 검증에서 최고 93.46%의 정확도 성능은 본 실험의 반려견 데이터셋을 제공한 기욤 등[4]이 기록한 92%의 성능을 능가하는 결과이며, 임계값의 세밀한 조정을 통해 추가적인 성능 향상을 얻을 수 있을 것으로 생각된다. 응용 서비스에 대한 사용자 인터페이스를 적

표 4. ResNet-50 기반의 실험 결과 정리
Table 4. Summary of ResNet-50 based experiments

Items		Exp. 5 (Pretrained=false)	Exp. 6	Exp. 7	Exp. 8
Common values	Number of epochs	200			
	Dataset split	Train:Test = 7:3			
	Learning rate	1.00E-04			
	Dropout rate	0.3			
Hyper-parameters	Hard triplet ratio	0.2	0.2	0.2	0.2
	Triplet margin	0.5	0.6	0.3	0.5
	Threshold for verification	1.2	1.2	0.8	1.2
Accuracy results (%)	Verification	90.47	92.00	93.46	93.23
	Recognition_rank1	60.96	81.15	77.49	80.16
	Recognition_rank5	78.80	91.16	90.8	91.44

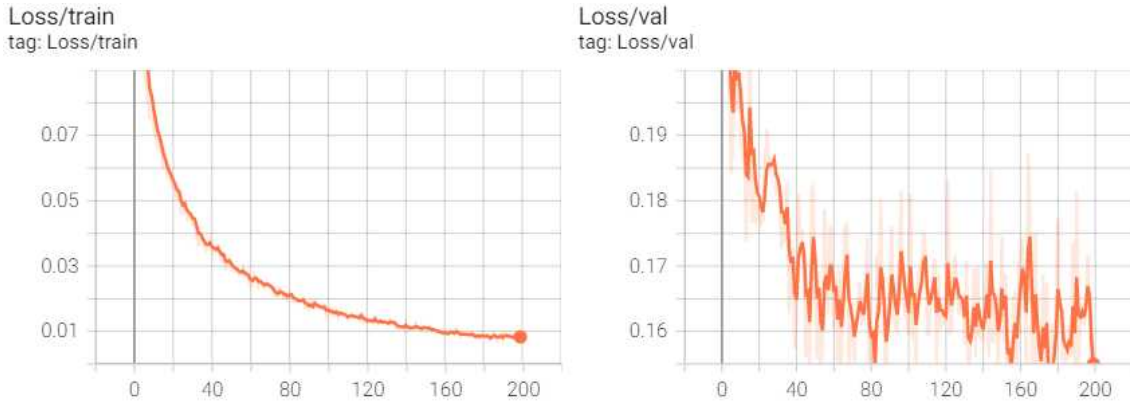


그림 11. 실험 8에 대한 손실값의 변화
 Fig. 11. Loss values variation for Exp. 8

절히 구성할 경우, 안면 식별 대신, 보다 성능이 좋은 안면 검증 기능만으로도 서비스 효과를 얻을 수가 있다.

그림 11은 표 4에서 안면 식별 정확도에서 가장 좋은 성능을 얻은 실험 8에 대한 학습시 손실값의 변화를 나타낸다.

V. 결론 및 향후 과제

2021년 동물보호법의 개정에 따라 반려동물에 대한 등록 의무화가 시행되었음에도 불구하고, 내장형 칩 또는 외장형 태그로 이루어지는 등록 방법의 불편함과 비용 등으로 등록율이 낮은 실정이다. 본 논문에서는 이에 대한 대안으로 새로운 반려견 등록 방법으로 고려되고 있는 반려견 인식 방법에 대해서, 딥러닝 학습을 통해 반려견의 개체 인식 모델을 만들고, 다양한 방법으로 실험을 통해 인식 성능을 개선하고자 하였다.

본 실험에서는 1,393마리의 반려견 사진 8,360장으로 기본 데이터로 구성하고, 컬러 지터, 수평 플립, 랜덤 로테이션 등의 데이터 증강 기술을 사용하여 실험 데이터셋을 구성하였다. 현재까지 사람의 얼굴 인식 등에서 가장 많이 이용되고 있는 InceptionNet 모델과 ResNet 모델을 백본 네트워크로 사용하였다.

얼굴 인식에서 가장 뛰어난 손실함수인 삼중항 손실 방법으로 학습하였으며, 이 학습 방법에서 중요한 파라미터

인 삼중항 마진, hard triplet 비율의 선택, 안면 검증 임계값 등을 변화하면서 최적의 학습 효과 요소를 찾아 실험하였다. 안면 검증(verification)과 안면 식별(recognition)로 나눠 얼굴 인식 실험을 수행한 결과, ResNet-50 기반의 모델에서 최고 93.46%의 안면 검증 정확도 성능을 얻을 수 있었다. 안면 식별 시험에서는 rank-1에서 81.15%, rank-5에서 91.44%의 최고 성능을 각각 얻을 수 있었으며, 이는 국내의 기존 연구에서 보여준 성능과 유사하거나 더 우수한 결과이다.

본 실험과정에서 삼중항 손실로 학습할 경우에 적절한 hard triplet 비율을 확인할 수 있었으며, 삼중항 마진값과 안면 검증 임계값이 성능에 중요한 영향을 미치는 것을 확인하였다. 좋은 성능을 위해 가장 중요한 요소는 양질의 학습 이미지를 대량으로 확보하는 것이다. 본 실험과정에서 데이터셋의 규모를 늘리기 위해 (사)반려동물협회에서 취득한 추가 이미지 데이터로 실험하였으나, 노이즈가 많은 이미지의 경우 얼굴 부분 검출이 잘 되지 않았다. 대규모 데이터셋의 확보는 추후 지속적으로 수행할 과제이다.

국내에서 지속적으로 확대되는 반려견 산업에 맞추어, 이와 관련한 데이터의 체계적인 구축과 딥러닝 기술을 이용한 인식 방법 연구 등은 더욱 필요하다고 생각된다. 특히 동물보호센터에서 보다 정확한 유기·유실견 조기 탐지를 통해서 소유주에게로의 복귀율을 높이고, 정부의 반려견 개체 관리 분야에서도 활용도가 높을 것으로 기대된다.

참 고 문 헌 (References)

- [1] D. Lee and G. Park, "Analysis of Transfer Learning Effect for Automatic Dog Breed Classification," *Journal of Broadcasting Engineering*, Vol. 27, No. 1, pp.133-145, Jan. 2022.
doi : <https://doi.org/10.5909/JBE.2022.27.1.133>
- [2] Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, "Dog Breed Classification using Part Localization," *European Conference on Computer Vision(ECCV)*, pp.172 - 185, 2012.
doi : https://doi.org/10.1007/978-3-642-33718-5_13
- [3] X. Wang, V. Ly, S. Sorensen, and C. Kambhamettu, "Dog breed classification via landmarks," *International Conference on Image Procening(ICIP) 2014*, pp. 5237 - 5241, January 2015.
doi : <https://doi.org/10.1109/ICIP.2014.7026060>
- [4] G. Mougeot, D. Li, and S Jia, "A Deep Learning Approach for Dog Face Verification and Recognition," *PRICAI 2019: Trends in Artificial Intelligence*, pp. 418 - 430, 2019.
doi : https://doi.org/10.1007/978-3-030-29894-4_34
- [5] S. Kumar and S. K. Singh, "Monitoring of pet animal in smart cities using animal biometrics," *Future Generation Computer Systems*, Vol. 83, pp. 553 - 563, 2018.
- [6] The DogFaceNet Dataset 1, <https://github.com/GuillaumeMougeot/DogFaceNet/releases> (accessed Mar. 02, 2022)
doi : <https://doi.org/10.1016/j.future.2016.12.006>
- [7] D. G. Lowe, "Distinctive Image Features from ScaleInvariant Keypoints," *International Journal of Computer Vision*, vol.60, issue.2, pp.91~110, 2004.
doi : <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [8] H. Bay, T. Tuytelaars, and L. Gool, "SURF: Speeded Up Robust Features," *Computer Vision and Image Understanding*, Vol.110, No.3, pp.346~359, 2008.
doi : https://doi.org/10.1007/11744023_32
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, Vol. 2, pp.1097 - 1105. 2012.
doi : <https://doi.org/10.1145/3065386>
- [10] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the Gap to Human-level Performance in Face Verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701 - 1708, 2014.
- [11] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815 - 823, 2015.
- [12] M. D. Zeiler and R. Fergus. "Visualizing and Understanding Convolutional Networks," *European Conference on Computer Vision(ECCV)*, pp.818-833, 2014.
doi : https://doi.org/10.1007/978-3-319-10590-1_53
- [13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going Deeper with Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9, 2015.
- [14] Y. Sun, X. Wang, and X. Tang. "Deeply Learned Face Representations are Sparse, Selective, and Robust," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2892-2900, 2015.
- [15] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. "SphereFace: Deep Hypersphere Embedding for Face Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.212-220, 2017.
- [16] H. Wang, Y. Wang, Z. Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salk Lake City, US, pp. 5265-5274, 2018.
- [17] M. Lee, J. Park, and J. Jeong, "An improved system of Dog Identification based on Muzzle Pattern," *Proceedings of the Korean Society of Broadcast and Media Engineers Conference*, pp.199-202, 2015.
- [18] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, "Dog Breed Classification using Part Localization," *European Conference on Computer Vision(ECCV)*, pp.172 - 185, 2012.
doi : https://doi.org/10.1007/978-3-642-33718-5_13
- [19] Adam Klein. *Pet Cat Face Verification and Identification*, Stanford University, CS230 Fall 2019.

저 자 소 개



이 동 수

- 1990년 : 경북대학교 전자공학과 공학사
- 1992년 : 경북대학교 전자공학과 공학석사
- 1994년 ~ 1999년 : SKT 기술부
- 2015년 ~ 현재 : 유한대학교 정보통신공학과 교수
- 2019년 ~ 현재 : 서울과학기술대학교 NID융합대학원 박사과정
- ORCID : <https://orcid.org/0000-0001-8961-4321>
- 주관심분야 : 머신러닝, 사물인터넷



박 구 만

- 1984년 : 한국항공대학교 전자공학과 공학사
- 1986년 : 연세대학교 대학원 전자공학과 석사
- 1991년 : 연세대학교 대학원 전자공학과 박사
- 1991년 ~ 1996년 : 삼성전자 신호처리연구소 선임연구원
- 1999년 ~ 현재 : 서울과학기술대학교 전자IT미디어공학과 교수
- 2006년 ~ 2007년 : Georgia Institute of Technology, Dept. of ECE. Visiting Scholar
- 2016년 ~ 2017년 : 서울과학기술대학교 나노IT디자인융합대학원 원장
- 2020년 ~ 현재 : 한국방송미디어공학회 부회장
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터비전, 실감미디어, 딥뉴럴네트워크