

일반논문 (Regular Paper)

방송공학회논문지 제27권 제1호, 2022년 1월 (JBE Vol.27, No.1, January 2022)

<https://doi.org/10.5909/JBE.2022.27.1.133>

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

## 반려견 자동 품종 분류를 위한 전이학습 효과 분석

이 동 수<sup>a)</sup>, 박 구 만<sup>a)†</sup>

### Analysis of Transfer Learning Effect for Automatic Dog Breed Classification

Dongsu Lee<sup>a)</sup> and Gooman Park<sup>a)†</sup>

#### 요 약

국내에서 지속적으로 증가하는 반려견 인구 및 산업 규모에 비해 이와 관련한 데이터의 체계적인 분석이나 품종 분류 방법 연구 등은 매우 부족한 실정이다. 본 논문에서는 국내에서 양육되는 반려견의 주요 14개 품종에 대해 딥러닝 기술을 이용한 자동 품종 분류 방법을 수행하였다. 이를 위해 먼저 딥러닝 학습을 위한 반려견 이미지를 수집하고 데이터셋을 구축하였으며, VGG-16 및 Resnet-34를 백본 네트워크로 사용하는 전이학습을 각각 수행하여 품종 분류 알고리즘을 만들었다. 반려견 이미지에 대한 두 모델의 전이학습 효과를 확인하기 위해, Pre-trained 가중치를 사용한 것과 가중치를 업데이트하는 실험을 수행하여 비교하였으며, VGG-16 기반으로 fine tuning을 수행했을 때, 최종 모델에서 Top 1 정확도는 약 89%, Top 3 정확도는 약 94%의 정확도 성능을 각각 얻을 수 있었다. 본 논문에서 제안하는 국내의 주요 반려견 품종 분류 방법 및 데이터 구축은 동물보호센터에서의 유기·유실견 품종 구분이나 사료 산업체에서의 활용 등 여러가지 응용 목적으로도 활용될 수 있는 가능성을 가지고 있다.

#### Abstract

Compared to the continuously increasing dog population and industry size in Korea, systematic analysis of related data and research on breed classification methods are very insufficient. In this paper, an automatic breed classification method is proposed using deep learning technology for 14 major dog breeds domestically raised. To do this, dog images are collected for deep learning training and a dataset is built, and a breed classification algorithm is created by performing transfer learning based on VGG-16 and Resnet-34 as backbone networks. In order to check the transfer learning effect of the two models on dog images, we compared the use of pre-trained weights and the experiment of updating the weights. When fine tuning was performed based on VGG-16 backbone network, in the final model, the accuracy of Top 1 was about 89% and that of Top 3 was about 94%, respectively. The domestic dog breed classification method and data construction proposed in this paper have the potential to be used for various application purposes, such as classification of abandoned and lost dog breeds in animal protection centers or utilization in pet-feed industry.

Keyword : Deep Learning, Transfer Learning, Resnet, VGGNet, Dog Breed

## 1. 서론

농림축산식품부의 2020년 반려동물 보호복지 실태조사<sup>[1]</sup>에 따르면, 국내에서 양육되는 반려견 수는 2020년 12월 등록 기준으로 약 232만여마리에 달하며, 그 숫자는 최근 5년간 연평균 증가율이 약 21.3%에 이를 정도로 지속적으로 증가하고 있다. 또한, 미등록된 반려견 수를 약 400만 마리로 추정하고 있다. 반려견 수의 증가와 함께 유기·유실견 수도 해마다 증가하고 있으며, 표 1에 나낸 것처럼 이를 관리하는 동물보호센터의 보호관리 비용도 2020년 한해에만 약 267억원으로 매년 증가하고 있는 추세이다.

표 1. 유실동물 구조 보호 비용

Table 1. Costs of Rescue and Care for Lost Pets

Year	2017	2018	2019	2020	CAGR
Amount (Unit: 100Mwon)	156	200.4	232	267.1	19.6%

현재 유기·유실견에 대한 공식적인 관리는 농림축산식품부가 각 지방자치단체와 연계하여 운영하는 동물보호관리 시스템을 통해 이루어지고 있다. 이 시스템의 기능중 하나로, 정부로부터 승인받은 전국 280개의 동물보호센터에서 유기·유실견에 대한 정보를 제공하고 있으며, 그림 1에서 보는 것처럼 한장의 사진과 함께 품종, 성별, 발견장소 등 간단한 내용을 정리하여 등록하고 있다.

하지만, 담당자가 수동으로 지정하는 품종 구분이 모호하여 대부분 믹스견(잡종견)으로 분류되는 문제가 있으며, 지역별 통계 데이터나 보호관리적 데이터는 있으나, 유실견 자체에 대한 데이터 구축은 부재한 상태이다. 유실견에 대한 조기 탐색이나 반려견 사료 산업의 발전 등 반려견에 대한 데이터의 체계화와 이를 이용한 산업적 응용은 필수적인 사항이 되고 있으나, 국내 반려견에 대해 활용가능한 데이터는 매우 제한적이라 할 수 있다. 반려동물 보호복지

실태조사에 따르면, 유기·유실견에 대한 정상적인 복귀(소유주 인도 또는 분양) 비율은 약 40%에 불과하며, 나머지는 안락사, 자연사 등으로 처리된다.

본 논문에서는 국내에서 양육되는 175종의 반려견 품종 중에서 상위 90% 이상을 차지하는 14개 품종에 대해, 반려견 자동 품종 분류 방법으로 딥러닝 알고리즘을 이용한 전이학습을 적용하고 그 효과를 분석해 보고자 한다. 동물보호관리시스템에서 운영되는 동물보호센터의 정보와 본 연구 결과물인 품종 분류 알고리즘을 연동함으로써, 유기·유실견에 대한 자동 품종 분류 기능을 추가하여 그들의 정상적인 복귀율을 높일 수 있을 것으로 기대한다.

국내에서 머신러닝 또는 딥러닝 기술을 활용하여 반려견 또는 반려묘에 대한 품종 분류 방법을 연구한 사례는 찾아보기 어렵다. SSD(Single Shot Multibox Detector)와 YOLO(You Only Look Once) 등 딥러닝 객체 검출 모델을 이용하여 이미지로부터 반려견의 위치를 검출하는 연구<sup>[2]</sup> 등은 반려견과 다른 객체를 판별한다는 의미에서, 동일한 반려견 클래스 내에서 개의 품종을 구별해야 하는 서브클래스 분류 기술과는 차이가 있다고 할 수 있다.

딥러닝 기술이 적용되기 이전에 전통적인 컴퓨터 비전 방법으로 동물의 코에서 발견할 수 있는 고유한 비문(nose print)를 판별하여 개체를 인식하는 연구가 몇차례 시도되었다. 대표적으로 국소 특징점 추출방식의 하나인 SURF(Speeded Up Robust Feature)와 RANSAC(RANDOM Sampling Consensus) 알고리즘을 활용하여, 개의 비문을 비교함으로써, 개체를 인식하는 연구가 있었으나<sup>[3]</sup>, 이 방식은 적절한 이미지 획득이 매우 어렵고 정확도 성능 확보 측면에서 범용적 개체 인식 방법으로 활용하기가 어렵다.

본 연구에서는 먼저 국내의 주요 반려견 14개 품종에 대해 웹 크롤링과 온라인 반려동물 등록시스템으로부터 3,420장의 이미지를 확보하고 학습 데이터셋을 구성하였다. 또한 가중치를 업데이트하기 위한 전이학습(transfer learning)을 위해 J. Liu 등<sup>[4]</sup>이 수집한 8,351장의 반려견 이미지를 사용하였다. 전이학습을 위한 백본 네트워크로는 본 연구의 데이터 규모와 분류 레벨을 고려하여 VGG-16<sup>[5]</sup>과 Resnet-34<sup>[6]</sup> 모델을 사용하여, 실험을 진행하였다.

실험에서는 먼저 8,351장의 반려견 이미지 데이터셋을 두 백본 네트워크 모델에 적용하여 각 계층의 가중치를 업

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 received December 8, 2021; Revised December 14, 2021; Accepted December 14, 2021.



그림 1. 동물보호센터의 유실견 정보  
Fig. 1. Lost Dog Information on the Website of Pet Care Center

데이트한 후, 그 가중치를 활용하여 국내 반려견 이미지 3,420장의 이미지를 입력 데이터로 사용하고, 최종 출력단에서 14종으로 분류를 시도하였다. 두 백본 네트워크 중에서는 VGG-16 모델을 사용하였을 때, 보다 더 우수한 분류 성능을 얻을 수 있었다. 오픈 테스트 데이터 684장으로 분류 성능을 실험한 결과, Top 1 정확도로 88.59%를 기록하였으며, Top 3 정확도는 94.06%를 각각 보여주었다. Top 1은 분류 결과에서 최상위 확률값을 가지는 1개 결과만 취하는 것이며, Top 3는 상위 확률값을 가지는 3개 결과까지 포함하는 것을 나타낸다. 현재 동물보호센터에서 믹스견으로만 분류되는 경우가 많은 상황에서, 유실견 보호자에게 보여주는 데이터로서 Top 3까지의 분류 결과도 충분히 의미 있다고 할 수 있다.

본 논문 2장에서는 먼저 머신러닝 또는 딥러닝 기술을 적용하여 안면인식을 연구한 주요 사례들을 살펴보고, 3장에서는 반려견에 대한 데이터셋 구축 방법, 4장에서는 품종 분류를 위한 방법과 실험 내용을 기술하였으며, 5장에서 결론과 향후 계획에 대해 기술하였다.

## II. 관련 연구

### 1. 관련 연구동향

컴퓨터 비전 분야에서 2D 이미지로부터 목표하는 객체에 대해 특징을 추출하는 문제는 매우 중요한 이슈로 다루

어져왔다. 특히 얼굴 인식 기술은 딥러닝 기술을 적용하기 전까지도 전통적인 주성분분석(PCA: Principal Component Analysis), 선형판별분석(LDA: Linear Discriminant Analysis) 등 홀리스틱 기법(holistic methods)<sup>[7][8]</sup>를 거쳐, 국소 특징 추출(local feature extraction) 기술을 홀리스틱 기법과 합성하여 성능을 개선한 하이브리드 기법 등이 많이 연구되었다<sup>[9][10]</sup>.

하지만, 인간의 얼굴 인식과 달리 동물의 안면 인식은 매우 어려운 작업이며, 반려견의 경우 품종간 또는 개체간 매우 미묘한 차이를 가지고 있고 적절한 데이터셋 확보도 어렵기 때문에 이에 대한 연구는 매우 적으며, 성능 측정도 제약 조건을 가진 환경에서 이루어지고 있는 상황이다.

### 2. 딥러닝을 이용한 이미지 인식

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, 컴퓨터 비전에서 인간 수준에 가까운 성능을 보여주었다<sup>[11]</sup>. DeepFace 기술로 명명한 이 시스템에서는 페이스북에서 수집된 440만장의 대규모 얼굴 이미지를 사용하였다(SFC: Social Face Classification 데이터셋). 그림 2와 같이 딥러닝 학습을 위해 얼굴에 67개의 기준점을 설정하고 3D 정렬 기술로 정면화(frontalization) 전처리를 진행한 후, 딥러닝 네트워크에 넣어 얼굴 인식을 수행하였다.

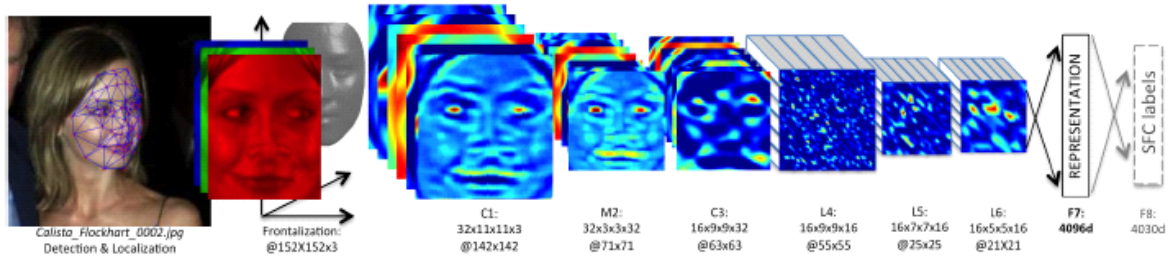


그림 2. DeepFace 시스템 아키텍처 요약

Fig. 2. Outline of the DeepFace Network Architecture

이 논문에서는 SFC 외에도 LFW(Labeled Faces in the Wild), YTF(YouTube Faces) 데이터셋을 이용하여 다양한 형태로 실험한 결과를 비교하여 보여주고 있다. LFW 기준 얼굴 비교검증(face verification)에서 97.35%의 얼굴 인식 정확도를 달성하였고, YTF 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<sup>[12]</sup>. 삼중항 손실은 그림 3에 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.

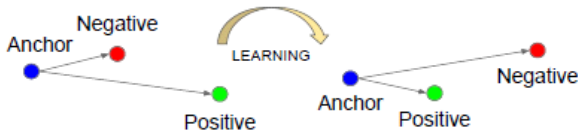


그림 3. 삼중항 손실 학습

Fig. 3. The Triplet Loss Learning

여기서,  $f(x_i^a)$ ,  $f(x_i^p)$ ,  $f(x_i^n)$ 는 각각 anchor 이미지, pos-

itive 이미지, negative 이미지에 대한 벡터 값을 나타내며,  $\alpha$  값은 서로 다른 클래스 간 차이를 만들기 위한 마진(margin)을 나타낸다.

이 논문에서는 약 8백만명의 사람들에 대한 사진 2억여 장을 학습 데이터셋으로 사용하였으며, 학습 데이터셋에 대해 다양한 크기의 크기변환 이외의 변로 전처리를 사용하지 않았다. 백본 네트워크로는 Zeiler&Fergus Net<sup>[13]</sup> 및 Inception Net<sup>[14]</sup> 기반 CNN 네트워크 4종류를 사용하여 실험하였다. CNN 네트워크에서는 각 이미지에 대한 128차원 임베딩 벡터를 추출한 후, 삼중항 손실로 학습하는 방법을 수행하였다. 벤치마크 데이터셋인 LFW에서 실험한 결과, 당시 SOTA(State-of-the-Art)였던 DeepID2+<sup>[15]</sup> 보다 오류율을 30%까지 줄이면서 1:1 얼굴 비교 검증(face verification)에서 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<sup>[16]</sup>나 Cosface 방법<sup>[17]</sup> 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.

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

### 3. 동물에 대한 생체 인식

사람의 얼굴 인식에 대한 연구 성과와는 달리, 국내외에서 반려동물에 대한 생체 인식을 통해 개체 또는 특정 품종을 식별하는 연구는 매우 적은 편이다.

해외에서는 초기연구로서, 콜롬비아 대학의 J. Liu 등이 개의 품종별로 얼굴 모형의 기하학적 모델을 구축하여 품종별 분류를 시도했다<sup>[4]</sup>. 그들은 Flickr, Google 사이트 등에서 총 133 품종의 개 사진 8,351장을 수집하여 데이터셋을 구성한 후, 그림 4에 보는 것처럼 SIFT(Scale-Invariant Feature Transform) descriptor를 사용하여 8개의 포인트로 외형 특징을 추출하였다. 분류에는 SVM(Support Vector Machine) 기반의 디텍터를 설계하였으며, 67%의 인식 정확도를 보여주었다.

X. Wang 등은 상기 논문을 개선한 방법을 제안하였다<sup>[18]</sup>. 이들은 개의 얼굴로부터 8개의 랜드마크를 기하학적으로 연결했을 때 품종별로 더 잘 구별할 수 있음을 발견하였다. 그림 5와 같이 랜드마크로부터 만들어진 얼굴의 기하학적 정보를 Grassmann 매니폴드로 투영하는 방식으로 시스

템을 설계함으로써, 단순히 기하학적 정보를 사용하는 방법보다 성능을 20%이상 개선했다. 이 논문에서는 이전의 J. Liu 논문에서 수집한 품종별 데이터셋을 그대로 사용하고 분류방법도 SVM을 사용했으나, 특히 8개의 랜드마크별로 어떤 포인트가 품종 분류에 더 효과적인지를 실험하기도 했으며, 코보다는 눈의 상대적인 위치가 더 도움이 된다는 사실을 확인하였다.

딥러닝을 이용한 인식 작업의 경우, 국소 특징점을 추출하기 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 손실<sup>[12]</sup>을 사용하여 Resnet-like 모델과 VGG-like 모델에 적용하는 방법을 발표했다<sup>[19]</sup>. 이들은 485마리의 개 이미지 3,148 장의 비교적 적은 데이터셋으로 얼굴의 세 포인트를 기준으로 정렬하는 전처리를 수행한 후, pre-trained model의 계층 구조와 하이퍼파라미터들을 변경하면서 실험하였다. 한쌍의 사진을 비교하는 얼굴 비교 검증(verification)에서는 그림 6의 Resnet-like 모델

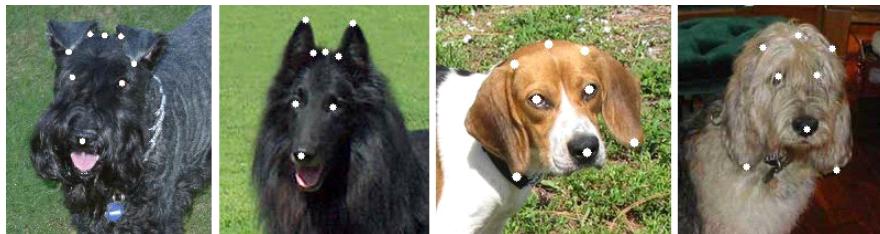


그림 4. 8개 부분으로 라벨된 이미지 샘플  
Fig. 4. Image samples labeled with 8 parts

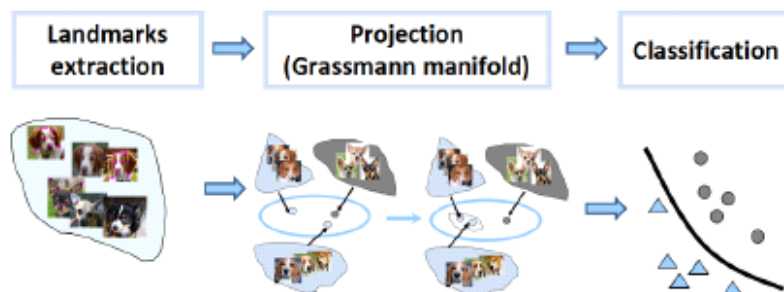


그림 5. Grassmann Manifold 투영방식의 분류 시스템  
Fig. 5. Classification System using Grassmann Manifold Projection



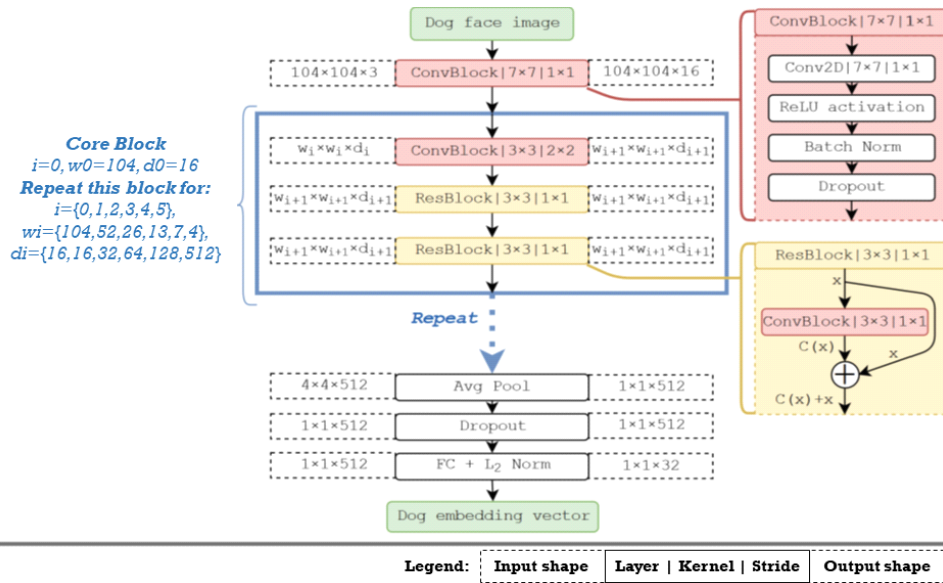


그림 6. Resnet-like 모델 정의  
 Fig. 6. Resnet-like Model Definition

로 최고 92%의 정확도를 달성하였고, 개체 식별(identification)에서는 Top 1에서 60.4%, Top 5에서 88.4%의 성능을 보여주었다. 딥러닝 기술을 개의 안면인식 응용에 적용했을 경우의 가능성과 어려움을 보여준 실험 결과로 상당히 의미있는 연구라 할 수 있다.

스탠포드 대학의 A. Klein은 딥러닝의 객체 검출 기술과 EfficientNet 등 몇가지 네트워크 모델을 이용하여 고양이 얼굴에 대한 인식 연구를 수행하였다<sup>[20]</sup>. 개체당 5장 이상의 사진을 가지는 고양이 사진 약 96,000장의 사진을 데이터셋으로 사용하였으며, 이전 연구와 달리 랜드마크 태깅이나 이미지 정렬 등의 전처리를 사용하지 않고, 그림 7에서 보는 것처럼 YOLOv3 객체 검출 알고리즘과 삼중한 손실을 EfficientNet에 적용하여 실험하였다. 얼굴 비교 검증에서 약 95%의 정확도로 SOTA 성능을 달성했으며, Top 5 개체 식별 성능으로 81.3% 성능을 보여주었다. 저자는

딥러닝 학습 방법론보다 고품질의 입력 데이터 DB가 있다면 더 좋은 성능을 가질 수 있다고 설명한다.

### III. 데이터셋 구축

국내에서 딥러닝에 적용하기 위해 필요한 반려견의 품종별 이미지 데이터셋을 구축한 사례는 찾아볼 수 없기 때문에, 먼저 국내 주요 반려견 14종에 대한 이미지 확보를 위해 먼저 구글 이미지 사이트에서 웹 크롤링을 통해 데이터를 수집하였다. 이와 별개로 반려동물보호시스템에 온라인 등록을 수행하는 “다움펫” 사이트(daumpet.com)에서 수집되는 이미지 데이터를 일부 추가하였다.

웹 크롤링을 통해 수집되는 이미지의 경우, 딥러닝 학습 데이터로는 유효하지 않는 노이즈 이미지가 많이 포함되어



그림 7. YOLOv3를 이용한 고양이 얼굴 검출  
 Fig. 7. Cat Face Detection using YOLOv3

있다. 예를 들면 그림 8에 보는 것처럼, 상품 광고 이미지, 모형 이미지, 캐릭터 이미지, 인형 이미지 등은 학습 데이터로 적절하지 않으므로 사전에 필터링 작업이 필요하다.



그림 8. 웹 크롤링으로 수집된 유용하지 않은 이미지 사례  
Fig. 8. Ineffective Images from Web Crawling

사전 필터링 작업을 거쳐 확보된 데이터는 총 3,420장이며, 14개 품종 이름과 품종별 유효 데이터 수를 표 2에 표시하였다.

본 실험에서는 표 3에 표시한 것처럼 두 종류의 데이터셋으로 구분하여 학습에 사용하였다. DS1은 J. Liu 등<sup>[19]</sup>이 논문에서 사용했던 133종 8,351장의 이미지로서, Americal Kennel Club(AKC)에서 인정한 반려견 품종 이미지이며 깃허브(github) 사이트를 통해 이용이 가능하다. DS1은 VGG-16 및 Resnet-34 네트워크의 가중치를 업데이트하기 위한 파인 튜닝(fine tuning) 전이학습을 수행하는 데 적용하였다.

DS2는 웹 크롤링과 동물등록시스템 사이트를 통해 수집된 국내 주요 반려견 품종 데이터셋이며, DS1을 통해 획득한 가중치 업데이트 정보를 두 기반 네트워크 모델에 적용할 때, 입력 데이터로 사용함으로써, 국내 반려견을 학습하고자 하는 본래의 목표에 초점을 맞추고자 하였다. 또한 DS2는 가중치가 업데이트되기 전의 pre-trained 값을 가지는 VGG-16 및 Resnet-34 모델에 적용하여 결과를 획득하고, 가중치를 업데이트했을 때와의 결과와 비교함으로써, 전이학습의 효과를 확인하고자 하였다.

DS1과 DS2 모두 전체 데이터는 학습 데이터와 테스트

표 2. 반려견 품종별 수집 유효 데이터수

Table 2. Effective Number of Images in Dog Breed Data Collection

No.	Breed Name (Label in Code)	Number of Images	No.	Breed Name (Label in Code)	Number of Images
1	Bichon Frise (bich)	323	8	Maltese (malz)	270
2	Bulldog (buld)	233	9	Pomeranian (pome)	283
3	Border Collie (coli)	274	10	Poodle (pudl)	227
4	Chihuahua (chua)	177	11	Schnauzer (schn)	249
5	Dachshund (dach)	257	12	Shih Tzu (sizu)	248
6	Golden Retriever (retr)	244	13	Welsh Corgi (wels)	182
7	Jindo (jind)	237	14	Yorkshire Terrier (yokt)	216
Total Number of Images: 3,420					

표 3. 실험에 사용된 데이터셋

Table 3. Datasets used in Experiments

Name	Source	Number of Images	Applied Network
DS1	Downloaded from Github	8,351	VGG-16, Resnet-34
DS2	Collected from web crawling	3,420	VGG-16, Resnet-34



그림 9. 품종 클래스별로 라벨링된 이미지 샘플

Fig. 9. Image Samples Classified with Breed Label

데이터 비율을 80:20으로 나누어 사용하였으며, 학습을 위한 배치 사이즈는 64로 설정하였다. 폴더별로 클래스를 분류하는 방식이므로 품종별 폴더 이름이 클래스 구별을 위

한 라벨이 된다. 그림 9는 각 클래스별로 분류된 학습 데이터셋에 대해 임의로 추출한 샘플을 보여준다. 새로 구축한 이미지 데이터의 사이즈는 모두 다르므로, VGG-16 및 Resnet-34 모델에 적용하기 위해 입력 이미지를 224x224 크기로 resizing 하였으며, 데이터 증강(augmentation) 기법으로는 random rotation을 사용하였다. 또한, RGB 채널별로 이미지의 각 픽셀값은 0~1 사이의 값으로 정규화되었으며, 각 에포크(epoch)마다 입력 데이터는 셔플링되어 입력 된다.

## IV. 딥러닝 모델 적용 실험

### 1. VGG-16 모델의 특징

VGGNet은 이미지 데이터 분류 문제에서 컨볼루션 네트워크 계층의 깊이(depth)가 성능에 어떤 영향을 주는지를

표 4. VGGNet의 ConvNet 구조

Table 4. ConvNet Configurations at VGGNet

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					



연구하기 위해 만들어진 모델이며, 심플한 구조와 우수한 성능으로 이후 많은 분류 문제에서 활용되어온 기술이다. 이 목적을 위해 VGGNet은 매우 작은 3x3 필터 사이즈만을 사용하고 스트라이드는 1로 설정하여, 네트워크 계층을 증가시키는 구조를 가진다. 표 4는 VGGNet에서 실험했던 네트워크 구성을 나타내며, A구조에서 E구조까지 볼드체로 표시된 컨볼루션 계층만을 증가시키면서, 저자들은 네트워크 계층이 깊어질수록 좋은 성능을 얻는다는 것을 확인하였다.

하지만 VGG-19에서는 VGG-16에 비해 더이상 성능이 개선되지는 않았으며, VGG-16에서 에러율이 조금 더 낮은 결과로 오히려 더 좋은 성능을 보여주었다. 따라서 본 논문에서도 VGG-16을 백본 네트워크로 사용하였다. 이 연구 결과를 보면, 네트워크의 깊이를 더 깊게 하면서 에러율을 개선하기 위해서는 단순히 컨볼루션 계층을 증가시키는 방법이 아닌 다른 네트워크 구조가 필요하다고 추측할 수 있다. VGGNet은 네트워크의 깊이에 의해 이미지 분류의 성능 개선이 가능함을 증명한 연구이며, Resnet 등에 의해 네트워크의 깊이가 대폭 증가하는 기반이 되었다.

## 2. Resnet-34 모델의 특징

Resnet의 네트워크 구조는 VGGNet 모델을 기반으로 하여, 컨볼루션 계층을 대폭 증가시켜 매우 깊은 신경망 구조를 가진다. 계층을 깊게 만들기 위해 저자들은 잔차 블록(residual block)을 고안하여 학습시킨다. 그림 11에 나타낸 것처럼 기존의 순차적인 입출력 구조와 달리, 네트워크의 입력 값을 복제하여 일정 크기의 컨볼루션 계층이 지날 때마다 출력 값에 더해주는 skip connection 구조를 가진다. 본 실험에 적용한 Resnet-34에서는 2개의 컨볼루션 계층 단위로 skip connection이 일어난다.

그림 10에서 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.

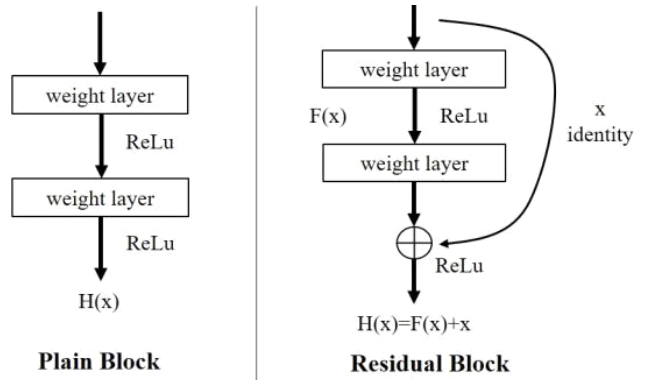


그림 10. Skip Connection을 가지는 잔차블럭

Fig. 10. Residual Block with Skip Connection

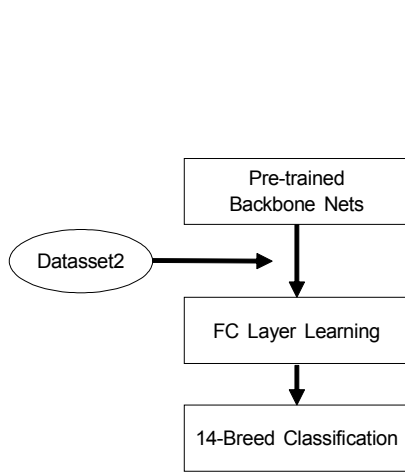
Resnet-34 구조에서는 입력층과 출력층을 제외하고 32개의 컨볼루션 신경망 계층을 가진다. 이전의 VGG 모델에서처럼 3x3 크기의 필터만 사용했으며, 계층이 진행되면서 특성 맵(feature map)의 크기가 줄어드는 것을 보완하기 위해, 신경망의 깊이를 64개, 128개, 256개, 512개로 점차 증가시킨 구조를 가진다.

## 3. 백본 네트워크 기반 전이학습 실험

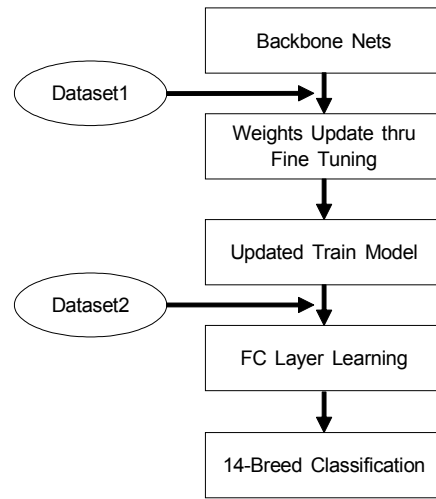
본 실험에서는 VGG-16과 Resnet-34를 기반으로 하여 pre-trained 가중치를 사용했을 경우와 새로 가중치를 업데이트했을 경우의 두 단계로 전이학습 방법을 수행하고 그 결과를 비교 분석하였다.

1단계 학습은 그림 11 (a)에 나타낸 것처럼 두 백본 네트워크에 대해서 각각 네트워크의 특징 추출 가중치를 그대로 활용하고, 출력단의 클래스는 14개로 조정하여 학습하는 방법이다. 입력 데이터만 반려견 이미지를 사용하였을 때 결과를 확인함으로써, 가중치를 업데이트한 전이학습을 적용했을 때의 결과와 비교해 보고자 하였다.

학습 데이터 수가 크지 않고 기존 학습 모델의 특성을 왜곡하지 않도록 하기 위해 비교적 낮은 학습률(learning rate) 0.001을 사용하였고, Adam 옵티마이저를 사용하였다. 손실함수는 다중 분류를 위한 크로스엔트로피(cross-entropy)로 사용하였으며, 배치 사이즈는 64, 에포크는 100번



(a) Classification by Feature Extraction



(b) Classification by Fine Tuning using Two Datasets

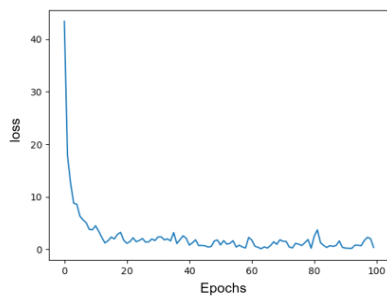
그림 11. 두 단계 전이학습

Fig. 11. Two-Stage Transfer Learning

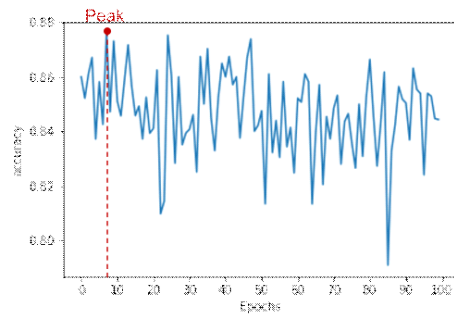
을 수행하였다. 학습 데이터셋은 14종 이미지 2,736장을 사용하였으며, 테스트 데이터셋은 684장의 이미지를 사용하

였다.

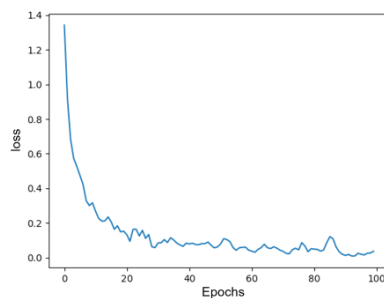
그림 12는 두 백본 네트워크에 대해 학습을 진행한 결과,



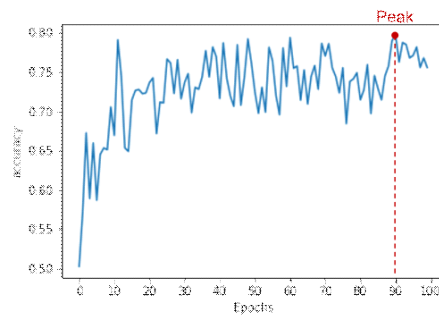
(a) Loss Value Variation: VGG-16



(b) Accuracy Variation: VGG-16



(c) Loss Value Variation: Resnet-34



(d) Accuracy Variation: Resnet-34

Fig. 12. 1단계 실험: 손실값 변화 및 정확도 변화

그림 12. 1st Stage Experiments: Variations of Loss Value and Accuracy

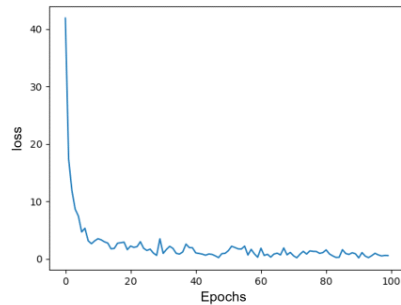
손실값의 변화와 분류 정확도 성능을 보여준다. 먼저 VGG-16 기반에서는 그림 12 (b)에서 보는 것처럼 100번의 에포크중 8번째에서 87.60%의 최고 성능치를 보여주며, 이후 정확도 성능은 지속적으로 감소하고 있어, pre-trained 모델만으로는 반려견의 특징값을 잘 추출한다고 보기는 어렵다. 그림 12 (d)는 Resnet-34 기반에서 동일한 학습 방법을 적용했을 경우 정확도 성능을 보여준다. 90번째 에포크에서 79.68%의 최고 성능을 얻어서, VGG-16 모델에 비해 정확도는 다소 낮아졌다.

2단계 전이학습은 그림 11 (b)에 나타난 것처럼 먼저 DS1 데이터를 사용하여 백본 네트워크의 가중치를 업데이트하는 파인 튜닝(fine tuning)을 수행한 후, 새로 획득한 가중치를 목표 데이터셋인 DS2를 적용하여 결과를 확인하는 방법이다. 즉, 기학습된 가중치를 사용하지 않고 133종 8,351장의 반려견 이미지에 대한 학습을 통해 새로운 가중치를 획득한 후, 최종적으로 DS2 데이터를 적용할 때에는 업데이트된 가중치를 동결(frozen)한 상태에서 출력단에서

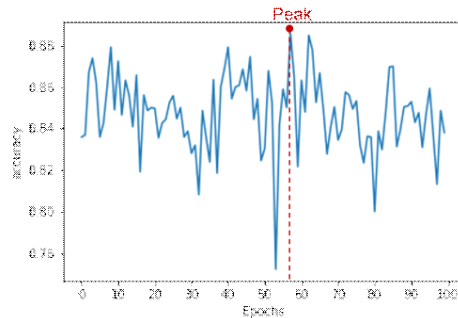
의 학습만 수행함으로써, 국내 반려견 14개 품종에 대한 분류 결과를 확인하였다.

2단계 실험에서의 하이퍼파라미터 설정과 학습 방법은 1단계에서 사용했던 것과 동일하게 적용하였다. VGGNet 논문에서와 같이, 1단계와 2단계 모두 검증(validation) 데이터로 테스트 데이터를 사용하였으며, 정확도가 최고치를 기록한 후 하락하는 추세를 확인하고 학습을 종료하였다.

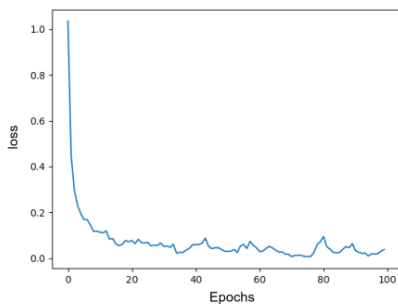
그림 13은 두 백본 네트워크에 대해 2단계 전이학습을 진행한 결과를 보여준다. 먼저, 2단계에서 에포크 진행에 따른 손실값의 변화를 보면, 이미 가중치가 업데이트된 모델을 사용하기 때문에 그림 13 (a) 및 (c)에서 보는 것처럼 1단계 실험보다 더 빨리 손실값이 줄어드는 모습을 볼 수 있다. 그림 13 (b)는 VGG-16 네트워크 기반의 분류 정확도 성능 추이이며, 최고 성능은 57번째 에포크에서 88.59%를 보여준다. 그림 13 (d)는 Resnet-34를 기반으로 한 성능 추이이며, 70번째 에포크에서 81.88%의 정확도로 최고 성능을 확인하였다.



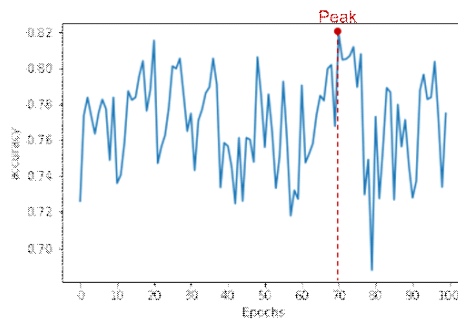
(a) Loss Value Variation: VGG-16



(b) Accuracy Variation: VGG-16



(c) Loss Value Variation: Resnet-34



(d) Accuracy Variation: Resnet-34

Fig. 13. 2단계 실험: 손실값 변화 및 정확도 변화

그림 13. 2nd Stage Experiments: Variations of Loss Value and Accuracy

표 5. 두 단계 실험결과 성능 요약

Table 5. Summary of Two-Stage Experiments

Backbone Network	Item	Peak Accuracy		Hyperparameters in Experiments
		1st Stage	2nd Stage	
VGG-16	Top 1	87.60%	88.59%	Learning Rate: 0.001 Batch Size: 64 Number of Epochs: 100 Optimizer: Adam
	Top 3	93.07%	94.06%	
Resnet-34	Top 1	79.68%	81.88%	
	Top 3	88.88%	90.65%	

본 실험 결과에 따른 전이학습의 효과를 보면, 두 백본 네트워크의 Feature extraction을 사용한 결과보다 가중치를 업데이트한 Fine tuning 전이학습의 효과가 크게 개선된다고 볼수는 없으나 약간 향상된 결과를 보여준다는 것을 알 수 있다.

서론에서 언급한 것처럼 본 연구의 목적중 하나인 유실견 탐색을 위한 수단으로 사용할 경우를 고려하여, 두 단계의 실험에서 각각 Top 3까지의 분류 정확도 성능을 확인해 보았다. <표 5>에서 각 단계별로 Top 1 및 Top 3에 대한 최고 성능값을 요약하였으며, 본 실험에서 사용한 주요 하이퍼파라미터 값을 표시하였다.

<표 5>에 나타난 것처럼, 2단계 전이학습 실험을 기준으로 Top 3까지의 최고 성능을 보면, Resnet-34에서는 90.65%를, VGG-16에서는 94.06%의 분류 정확도를 각각 보여줌으로써, 국내 반려견 품종 분류에 대한 유의미한 결과를 얻었다고 할 수 있다.

Top 3에서 94.06%의 성능을 보여줌으로써, 현장에 적용가능한 수준의 알고리즘을 얻을 수 있었다.

하지만 더 높은 성능을 얻기 위해서는 보다 다양하고 많은 유효 이미지 데이터 수집이 필요하다고 생각되며, 현재 SNS를 이용하여 추가적으로 반려동물의 이미지 데이터를 수집하는 작업을 진행하고 있다. 대규모 데이터 셋을 구축하게 되면 자동 품종 분류 뿐만 아니라 반려동물의 각 개체별 분류도 가능할 것으로 예상되어, 활용도가 대폭 늘어날 것으로 생각된다.

국내에서 지속적으로 증가하는 반려견 개체수 및 산업 규모를 고려할 때, 이와 관련한 데이터의 체계적인 분석이나 품종 및 개체 분류 방법 연구 등은 더욱 필요성이 커질 것으로 예상된다. 특히 동물보호센터에서 보다 정확한 유기·유실견 품종 분류를 통해서 보호자한테로의 복귀율을 높일 수 있을 것으로 기대되며, 향후 펫푸드 산업체에서의 활용 등 여러가지 목적으로 활용될 수 있는 가능성을 가지고 있다.

## V. 결론 및 향후 과제

본 논문에서는 국내에서 양육되는 주요 반려견 14개 품종에 대해, 딥러닝 알고리즘을 사용하여 반려견에 대한 자동 품종 분류 방법을 개발하고 성능 결과를 확인해 보았다. 이를 위해 웹 크롤링과 동물등록시스템으로부터 이미지 데이터를 수집한 후 작은 규모의 유효 데이터셋을 만들었다. 또한, 미국에서 구축된 딥러닝 학습용 반려견 데이터셋을 활용하여, VGG-16 및 Resnet-34 네트워크를 기반으로 하는 파인 튜닝 전이학습에 적용하였다. 이 과정을 통해 업데이트된 가중치를 딥러닝 모델에 적용하고, 국내 반려견 이미지 데이터셋을 적용하여 품종 분류를 실험한 결과, VGG-16을 기반으로 한 정확도 최고치로 Top 1에서 88.59%.

## 참 고 문 헌 (References)

- [1] Ministry of Agriculture, Food and Rural Affairs, Report on the Protection and Welfare Survey for Pets 2020, News Publishment, May 18, 2021.
- [2] J. Cho, C. Lee, M. Kim, S. Kim, and S. Jang, "Study of Pets Detection." *Proceedings of the Korean Institute of Information Scientists and Engineers Conference*, pp.1527-1529, 2020.
- [3] 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
- [4] 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.
- [5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556v6*, 2015.

- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770 - 778, 2016
- [7] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *Journal of the Optical Society of America A*, vol. 14, no. 8, pp. 1724 - 1733, 1997.
- [8] W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets, and J. Weng, *Face Recognition: From Theory to Applications*, pp. 73 - 85, Springer, 1998.
- [9] C. Chan, J. Kittler, and K. Messer, "Multi-scale Local Binary Pattern Histograms for Face Recognition," *International conference on biometrics*, pp. 809 - 818, Springer, 2007.
- [10] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based Methods in the Wild," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, 2008.
- [11] 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.
- [12] 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.
- [13] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *European Conference on Computer Vision(ECCV)*, pp.818-833, 2014.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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,
- [18] X. Wang, V. Ly, S. Sorensen, and C. Kambhamettu, "Dog Breed Classification via Landmarks," *IEEE International Conference on Image Processing (ICIP)*, pp. 5237 - 5241, 2015, <https://doi.org/10.1109/ICIP.2014.7026060>.
- [19] 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, [https://doi.org/10.1007/978-3-030-29894-4\\_34](https://doi.org/10.1007/978-3-030-29894-4_34).
- [20] 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 : 한국방송 미디어공학회 부회장
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터비전, 실감미디어, 딥뉴럴네트워크