

# 딥러닝을 활용한 반려견 탐지 및 행동 분류

문효원, 정희철\*

경북대학교 융합학부, \*경북대학교 인공지능학과

madit1023@knu.ac.kr, \* heechul@knu.ac.kr

## Dog Detection and Behavior Classification using Deep Learning

Hyowon Moon, Heechul Jung\*

Department of Convergence, Kyungpook National Univ.,

\* Department of Artificial Intelligence, Kyungpook National Univ.

### 요 약

집에 홀로 남는 반려동물의 관찰을 위해 홈 CCTV를 사용하는 비율이 늘어나고 있다. 하지만 이를 활용하여도 보호자가 직접 관찰하지 않으면 반려동물의 정보를 알 수 없다. 이를 해결하기 위해, 본 연구에서는 딥러닝을 활용한 반려동물 영상 분석 시스템을 제안한다. YOLOv5를 활용한 반려견 객체 탐지 모델과 CNN을 활용한 객체 및 행동 분류 모델을 구축하였다. 학습 결과, 객체 탐지 모델에서 92.5%, 객체 및 행동 분류 모델에서 객체 97.58%, 행동 90.48%의 정확도를 보였다. 이를 통해 영상에서 반려동물 정보를 제공하는 시스템에 대한 가능성을 확인하였다.

### I. 서 론

국내 반려동물 양육 가구는 604만 가구로, 전체 가구의 29.7%가 반려동물과 함께하고 있다 [1]. 반려동물 돌봄에 대한 관심도 높아지면서 집에 혼자 남겨지는 반려동물을 위한 스마트 기기들이 주목받고 있다. 그 중 홈 CCTV 사용이 큰 비중을 차지하고 있다. 하지만 CCTV를 이용해도 보호자가 항상 관찰하는 것은 불가능하기에 지속적으로 확인할 수 없다. 따라서 영상을 자동으로 분석하여 반려동물 정보를 제공하는 시스템이 필요하다.

본 논문은 YOLOv5 [2,3]를 이용하여 영상에서 반려견 객체를 탐지하는 모델과 CNN 계열 모델을 이용하여 객체와 행동을 분류하는 모델을 연구하였다. 객체 탐지에서는 모델 사이즈, 이미지 크기를 조정하여, 객체 및 행동 분류에서는 AlexNet [4], ResNet [5], ConvNeXt [6] 모델을 활용하여 실험하였다. 이때, 동일 건중 반려견 사진을 이용, 세밀한 분류를 수행하였다.

### II. 본 론

#### 가. 데이터 셋 구성

객체 탐지 모델의 이미지 데이터는 총 276장으로, 직접 촬영한 반려견 사진 192장, 크롤링으로 수집한 개 사진 84장으로 이루어져 있다. 라벨링은 Label Studio [7]를 활용, 반려견 3마리 개체 {Soonsim, Hana, Arong}와 다른 개체로 4가지 Class를 분류하였다.

객체 및 행동 분류 모델의 이미지 데이터는 총 1,242장으로, 직접 촬영한 반려견의 사진으로 이루어져 있다. 객체 분류와 행동 분류에 대해 각각의 데이터셋을 구성하였다. 객체 분류의 경우 3마리 반려견으로, 행동 분류의 경우 8개의 Class {보통, 졸음, 편안, 신남, 기다림, 섭취, 애정, 응시}로 분류하였다. 해당 분류는 개의 표정이 인간을 모방한다는 연구 [8]를 고려하여 선정하였다.

#### 나. 객체 탐지 모델

평가 기준은 mAP@0.5을 사용하였다. AP는 임계치 이상 IoU (Intersection over Union)일 경우를 정답으로 인정하여 P-R (Precision-Recall) 곡선을 그린 다음, 면적을 구한 것으로, mAP는 Class별 AP를 평균 낸 값이다. IoU 임계치 0.5에 대해 mAP를 구한 것을 mAP@0.5로 표기한다.

실험은 COCO 데이터셋에 pretrain된 weight를 전이 학습을 통해 미세 조정(fine-tuning)하는 방식으로 진행되었으며, 배치 사이즈 32, SGD(Stochastic gradient descent) optimizer, 학습률 0.01을 사용하였다.

#### 1) 실험 - 모델 사이즈 변경

YOLOv5의 경우 4가지 사이즈의 모델이 제공된다. 적절한 성능을 보장하면서 빠른 속도로 동작하는 모델을 찾기 위해, 30 epoch 동안 학습을 진행하였다.

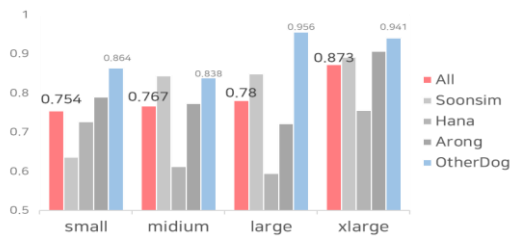


그림 1. 모델 사이즈에 따른 AP@0.5

실험 결과, Xlarge 모델이 87.3%로 가장 큰 값으로 나타났으며, 모델 사이즈가 클수록 성능이 좋았다. 이미지 한 장에 대한 연산 시간은 각각 약 12, 16, 20, 28ms로 나타났으며, 큰 모델에서도 빠른 탐지가 가능하다.

## 2) 실험 - 이미지 Resize 크기 변경

YOLOv5 모델에서, 입력되는 이미지의 사이즈를 변경하는 단계가 있는데, 이를 320, 512, 640로 조정하여 Xlarge 모델에서 30 epoch 동안 학습시켜 비교하였다.

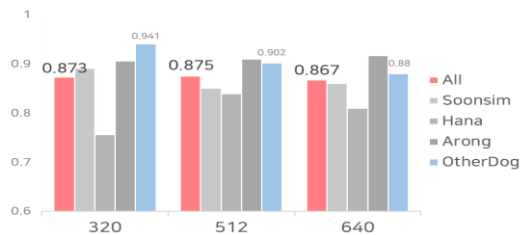


그림 2. resize 크기에 따른 AP@0.5

실험 결과, 동일 건중에 대해서는 이미지 사이즈가 커질수록 정확도가 높아졌고, 타 건중에 대해서는 정확도가 낮아졌다. 이는 이미지 사이즈가 커질수록 섬세한 특징에 대해서 잘 학습하나, 큰 부분의 특징에 대해서 학습하기 어렵기 때문으로 보인다.

## 3) 최종 모델

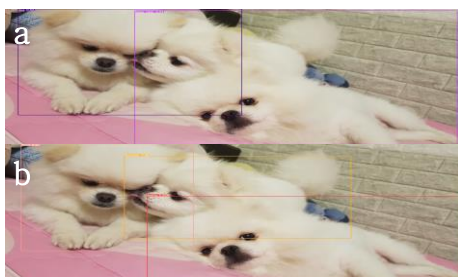


그림 3. (a) 학습 전, (b) 최종 모델

성능이 가장 좋았던 Xlarge 모델을 이미지 resize 크기 512로 100 epoch 학습시켰다. mAP@0.5 92.5%의 정확도를 보였다. 학습 전에는 붙어있는 반려견을 한 객체로 인식하거나, 곰, 인형으로 인식하는 경우가 있었다. 최종 모델은 붙어있는 반려견을 정확하게 인식했다.

## 다. 객체 및 행동 분류 모델

기존의 CNN 모델 중, AlexNet, ResNet18, ResNet50, ConvNeXt를 이용하여 실험을 진행하였다. ImageNet 데이터셋에 pretrain된 weight를 전이 학습을 통해 미세 조정하는 방식으로, 배치 사이즈 128, SGD optimizer, 학습률 0.0001로 100 epoch 동안 학습시켰다.

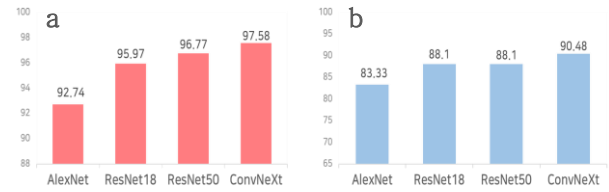


그림 4. (a) 객체 분류, (b) 행동 분류 결과

실험 결과, ConvNeXt 모델이 객체 분류 97.58%, 행동 분류 90.48%로 가장 높은 정확도를 보였다. 모델 사이즈가 클수록 정확도가 높았다.

## III. 결 론

본 논문에서는 pretrained weight를 fine-tuning하여 반려견 객체 인식 및 행동 인식에 적용할 수 있는 모델을 연구하였다. 모델의 학습용 데이터로 직접 촬영한 동일 건중 반려견 3마리의 사진과 크롤링한 사진을 이용하였다. 학습 결과 YOLO 기반 객체 탐지 모델 92.5%, CNN 기반 객체 분류 모델 97.58%, CNN 기반 행동 분류 모델 90.48%의 정확도를 보였다.

이후 연구를 통하여 발전시켜 나간다면, 반려동물 돌봄을 위한 도구로 활용될 수 있을 것이다. 일반 가정, 반려동물 유치원, 애견 호텔, 동물병원 등에서 위험 감지, 행동 패턴 분석, 학대 방지 등의 용도로 활용될 수 있다.

## ACKNOWLEDGMENT

“본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단(2020R1C1C100742311)의 지원을 받아 수행된 연구임.”

## 참 고 문 헌

- [1] KB금융지주 경영연구소, “2021 한국 반려동물보고서”, 2021, (<https://www.kbfg.com/kbresearch/report/reportView.do?reportId=2000160>).
- [2] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).
- [3] ultralytics, yolov5, 2020 (<https://github.com/ultralytics/yolov5>)
- [4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84–90.
- [5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [6] Liu, Zhuang, et al. "A convnet for the 2020s." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- [7] heartexlabs, label-studio, 2020 (<https://github.com/heartexlabs/label-studio/>)
- [8] Mota-Rojas, Daniel, et al. "Current advances in assessment of dog's emotions, facial expressions, and their use for clinical recognition of pain." Animals 11.11 (2021): 3334.