

## Deep Learning-Based Companion Animal Abnormal Behavior Detection Service Using Image and Sensor Data

Ji-Hoon Lee\*, Min-Chan Shin\*, Jun-Hee Park\*, Nam-Mee Moon\*\*

\*Student, School of Computer Science and Engineering, Hoseo University, Asan, Korea

\*Student, School of Computer Science and Engineering, Hoseo University, Asan, Korea

\*Student, School of Computer Science and Engineering, Hoseo University, Asan, Korea

\*\*Professor, Dept. of Computer Science and Engineering, Hoseo University, Asan, Korea

### [Abstract]

In this paper, we propose the Deep Learning-Based Companion Animal Abnormal Behavior Detection Service, which using video and sensor data. Due to the recent increase in households with companion animals, the pet tech industry with artificial intelligence is growing in the existing food and medical-oriented companion animal market. In this study, companion animal behavior was classified and abnormal behavior was detected based on a deep learning model using various data for health management of companion animals through artificial intelligence. Video data and sensor data of companion animals are collected using CCTV and the manufactured pet wearable device, and used as input data for the model. Image data was processed by combining the YOLO(You Only Look Once) model and DeepLabCut for extracting joint coordinates to detect companion animal objects for behavior classification. Also, in order to process sensor data, GAT(Graph Attention Network), which can identify the correlation and characteristics of each sensor, was used.

▶ **Key words:** Abnormal Behavior Detection, Behavior Pattern Analysis, Multimodal Analysis, Deep Learning, Wearable Device

### [요 약]

본 논문에서는 영상 데이터와 센서 데이터를 활용한 딥러닝 기반의 반려동물 이상행동 탐지 서비스를 제안한다. 최근 반려동물 보유 가구의 증가로 인해 기존 푸드 및 의료 중심의 반려동물 시장에서 인공지능을 더한 펫테크(Pet Tech) 산업이 성장하고 있다. 본 연구에서는 인공지능을 통한 반려동물의 건강관리를 위해 영상 및 센서 데이터를 활용한 딥러닝 모델을 기반으로 반려동물의 행동을 분류하고, 이상행동을 탐지하였다. 자택의 CCTV와 직접 제작한 펫 웨어러블 디바이스를 활용하여 반려동물의 영상 데이터 및 센서 데이터를 수집하고, 모델의 입력 데이터로 활용한다. 행동의 분류를 위해 본 연구에서는 반려동물의 객체를 검출하기 위한 YOLO(You Only Look Once) 모델과 관절 좌표를 추출하기 위한 DeepLabCut을 결합하여 영상 데이터를 처리하였고, 센서 데이터를 처리하기 위해 각 센서 별 연관관계 및 특징을 파악할 수 있는 GAT(Graph Attention Network)를 활용하였다.

▶ **주제어:** 이상행동 탐지, 행동패턴 분석, 멀티모달 분석, 딥 러닝, 웨어러블 디바이스

- First Author: Ji-Hoon Lee, Corresponding Author: Nam-Mee Moon
- \*Ji-Hoon Lee (monaminino@gmail.com), School of Computer Science and Engineering, Hoseo University
- \*Min-Chan Shin (shinmc0322@gmail.com), School of Computer Science and Engineering, Hoseo University
- \*Jun-Hee Park (cach456@gmail.com), School of Computer Science and Engineering, Hoseo University
- \*\*Nam-Mee Moon (mnm@hoseo.edu), Dept. of Computer Science and Engineering, Hoseo University
- Received: 2022. 08. 31, Revised: 2022. 10. 05, Accepted: 2022. 10. 12.

## I. Introduction

최근 급성장하는 고령화 및 출산율 감소와 같은 사회 구조적인 변화들에 따라 반려동물을 기르는 사람들의 수가 꾸준히 증가하고 있다. 이에 따라, 반려동물을 일간처럼 대하는 펫 휴머니제이션 현상이 가속화되고, 반려동물에게 제공되는 서비스의 품질이 인간이 누리는 서비스와 동일하게 제공되고 있다. 현재 제공되는 IoT(Internet of Things) 기반의 웨어러블 디바이스를 통한 헬스케어 서비스는 이러한 이유로 인해 대상이 인간에서 반려동물로 확장되고 있으며, 반려동물이 착용하는 웨어러블 디바이스에 관한 연구도 활발히 이루어지고 있다[1-3]. 반려동물 헬스케어 서비스는 사람과 반려동물 사이에서 의사소통의 어려움으로 인해 발생하는 문제들을 해소하기 위해 사전 진단을 통한 건강관리를 수행한다[4]. 이는 반려동물의 건강관리를 위한 행동 분석 연구들이 기반이 된다. 행동 분석 연구들은 대표적으로 영상과 센서 데이터를 사용하는 방법으로 나누어져 있다.

현재 센서 데이터를 이용한 행동 분석 연구들은 웨어러블 디바이스를 주로 사용하고 있다. 웨어러블 디바이스는 센서로부터 대상에 대한 세밀한 정보를 수집할 수 있으며, 수집된 센서 데이터는 행동 분석에 활용된다. 반려동물에 대한 센서 기반 행동 분석은 걷기, 뛰기, 앉기와 같은 행동들에 대한 다중적인 분류가 가능하다. 이러한 다중 행동 분류에서 센서 기반 행동 분석은 “걷는 행동”과 “뛰는 행동” 사이의 동적 행동에 정확한 분류가 가능하지만, 가만히 있는 정적 행동에 대해서는 분류 정확도가 떨어진대[5].

영상기반 행동 분석은 CCTV와 같은 실시간 영상 기반으로 객체 인식 및 행동 패턴을 분석하여 사건·사고 예방을 위한 연구가 주로 이루어지고 있다[6]. 또한, 반려동물 시장의 확대와 관심이 증감함으로써 분석의 대상이 동물로 확대되었다. 동물에 대한 영상기반 행동 분석은 “앉아 있는 행동”, “누워있는 행동”, “서있는 행동”과 같은 정적인 행동에 대해서는 높은 정확도의 성능을 보인다[7]. 그러나 “걷는 행동”과 “뛰는 행동”과 같은 세부적인 동적 행동 분류에 취약점을 가지고 있다.

멀티모달 분석은 감정, 사용자의 의도 그리고 행동 등을 분석하는데 있어서 영상, 음성, 텍스트 그리고 센서 데이터처럼 서로 다른 형태의 정보로 이루어져 뚜렷한 특성이 구분되는 데이터를 융합하여 사용한다. 멀티모달 분석은 서로 다른 형태의 데이터를 사용함으로써 각 데이터가 가진 단점을 보완하여 정확도를 개선하기 위한 분석 방법이다[8-9].

본 논문에서는 이러한 문제점을 해결하기 위해 멀티모달 반려동물 이상행동 탐지 모델을 제안한다. 센서가 부착된 아두이노와 고정형 CCTV를 통해 반려동물에 대한 영상 및 센서 데이터를 수집한다. 센서 데이터는 GAT(Graph Attention Network)를 사용하여 센서 간의 상관관계를 파악하고, 수집된 센서 값과 센서 간의 상관관계를 바탕으로 행동별 특징을 추출한다. 영상 데이터는 YOLOv4(You Only Look Once version 4) 모델을 통해 객체를 탐지하고, DeepSORT(SORT with A Deep Association Metric)를 적용하여 할당된 ID를 기반으로 객체를 추적한다. 마지막으로 반려동물의 추정 관절 좌표를 도출하기 위해 DeepLabCut 프레임워크를 적용한다. 추정 관절 좌표는 반려동물의 행동을 분류하기 위한 모델인 CNN(Convolutional Neural Networks)-BiLSTM(Bidirectional Long-Short Term Memory)의 입력 값으로 사용하여 정의한 행동으로 분류한다. 본 논문은 관련연구를 통해 각 데이터를 바탕으로 하는 행동 분류 방법에 대해 간략히 소개한다. 이후, 제안 모델에 대해 상세히 설명하며, 더불어 행동 분류를 위한 명확한 기준을 제시한다. 이후, 각 데이터의 처리 방법에 따른 실험 결과를 확인하며, 마지막으로 추후 발전 방향에 관해 논의한다.

## II. Related Works

동물의 이상행동을 탐지하기 위해 선행되어야 하는 연구로는 행동의 분류 연구가 있다. 행동 분류 연구가 선행되어야 예상치 못한 행동인 이상행동의 탐지가 가능하기 때문이다. 동물의 행동 분류는 다양한 데이터로부터 가능하며, 주로 영상 데이터와 센서 데이터를 기반으로 연구가 진행되었다.

### 2.1 Behavior classification based on video data

동물의 행동을 분류하고, 이상행동을 탐지하기 위한 가장 대표적인 데이터는 영상 데이터이다. 영상 데이터 기반 행동 분류 및 이상행동 탐지는 주로 다음과 같은 과정을 가진다. 객체 탐지, 객체 추적, 자세 추정, 행동 분류 및 이상행동 탐지 순서이다.

객체 탐지의 경우, 컴퓨터 비전 분야에서 주로 활용되는 기술이다. 이 기술은 주요 객체를 탐지하여 해당 객체를 중심으로 Bounding box를 표시하여 구분한다. 대표적인 방법으로는 YOLO가 있다[10]. YOLO는 단일 단계 방식을 가진 대표적인 알고리즘으로 원본 이미지를 같은 크기로 같은 크기의 그리드로 나눈다. 이후 각 그리드에 대해 그리드 중

양을 중심으로 미리 정의된 형태로 지정된 경계박스의 개수를 예측하고, 이를 기반으로 신뢰도를 계산한다. 미리 정의된 형태를 가진 경계박스의 수를 경계 상자라고 한다. 이와 특징을 기반으로 YOLO는 빠른 탐지 속도를 가지고 있으므로 실시간 영상에 적용될 수 있다는 장점이 있다[11-12].

객체 추적은 위의 객체 탐지 이후에 영상에서 특정 대상의 위치 변화를 추적하는 방법이다. 최근 딥러닝 기반 객체 탐지 기술을 통해 탐지된 대상에 관한 시공간적 특성의 유사성을 활용한 방법이 적용되고 있다. 객체 추적의 대표적인 방법으로는 DeepSORT가 있다. 이는 SORT(Simple Online and Realtime Tracking)의 문제점을 해결하기 위해 제안된 알고리즘이다[13]. SORT는 IoU(Intersection over Union)를 기반으로 물체에 대해 할당을 진행하므로 빠른 속도를 갖는 객체 추적이 가능하다[14]. 하지만, SORT 알고리즘은 기존의 대상을 잡지 못하고 다른 ID를 배정하는 ID Switching이 발생하는 단점을 가진다. 또한, 물체가 움직였을 경우와 대상이 가려지는 상황이 발생하였을 때 IoU가 작게 나와 결과가 무시되는 단점을 가진다. DeepSORT는 앞선 단점을 해결하기 위해 명백한 정보들을 통합하여 SORT의 성능을 향상하여, 대상이 장기간 가려진 경우에도 추적이 가능하고, ID Switching이 기존의 SORT 방식보다 45% 감소한 성능을 가지고 있다.

자세 추정은 사진이나 영상 속 객체의 신체 관절 구성을 바탕으로 위치를 추정하는 방법이다. 이를 위해 객체의 신체 윤곽이나 특정 신체 부위를 좌표와 같은 형태로 추출한다. 최근 이러한 신체 부위 예측의 정확도 및 효율을 높이기 위한 딥러닝 기반 자세 추정 프레임워크가 제안되고 있다. 대표적인 프레임워크는 DeeperCut과 DeepLabCut이 있다[15,16]. 그 중 DeepLabCut은 다양한 작업을 수행하는 동물의 마커리스 자세 추정을 위한 프레임워크이며, 영상의 프레임 기반으로 특징점의 움직임을 추정하는데 특화되어 있다. 또한, 최소 프레임의 훈련 데이터를 사용하여 심층 신경망을 사용한 전이 학습 기반으로 단일 및 복수 대상의 자세 추정을 위한 효율적인 방법을 제시한다. 이를 기반으로 기존에 사람 이외의 자세 추정이 힘들었던 DeeperCut의 한계점을 개선하여 다양한 대상에게서 움직임 추정이 가능하다.

마지막으로 행동 분류는 위와 같은 과정을 총합하여 움직이는 객체의 관절 좌표 데이터를 기반으로 행동을 정의하고, 정의된 행동을 통해 행동 분류 기준을 설정한다. 설정된 분류 기준을 통해 행동 좌표 데이터는 학습을 위한 데이터로 구성되어 기계학습 알고리즘을 통해 행동을 분류한다. 그러나 크기가 큰 데이터 셋의 경우 학습 속도가

느리고, 수치 특징이 많은 데이터일 경우 학습에 이상적이지 않은 단점을 가지고 있기 때문에 CNN과 BiLSTM을 사용한 연구가 진행되고 있다[17-18].

## 2.2 Behavior classification based on sensor data

동물에게 웨어러블 디바이스 등 센서 디바이스를 착용시켜 데이터를 수집하고, 행동을 분류하는 연구 또한 활발히 진행되었다[5][19-21]. 센서 데이터를 통한 행동 분류는 데이터의 수집 이후, 결측치 제거 및 이상치 제거 과정을 거친다. 이후 다양한 방법을 활용하여 행동 분류가 진행된다. SVM(Support Vector Machine) 또는 K-means 방식, 그리고 의사 결정 트리를 활용한 행동 분류가 주를 이루었다[5,22,23]. 특히 Reza Arablouei는 웨어러블 디바이스의 가속도 센서를 통해 방목된 소의 행동을 분류하였는데, 제안된 모델로는 컨볼루션 신경망으로도 볼 수 있는 유한 임펄스 응답(FIR, Finite Impulse Response)와 무한 임펄스 응답(IIR, Infinite Impulse Response)을 활용한 특징 추출 및 행동 분류 모델이 대표적이다[19]. 하지만, 행동 분류를 위해 가장 활발히 활용되는 의사 결정 트리는 문제 대부분에 잘 실행되는 범용성을 가지고, 중요하지 않은 특징은 제외하고 데이터 셋의 크기에 상관없이 모두 사용될 수 있다는 장점이 있으나, 트리 모델이 레벨 수가 많은 특징의 분할로 편향될 수 있으며 그로 인해 모델의 과적합 및 과소 적합이 발생하기 쉽다는 단점이 있다.

본 논문은 위와 같은 이유로 이전 연구에 활용한 GAT를 활용하고, 이를 영상 데이터와 결합하여 이상행동 탐지 서비스를 제안한다. GAT는 Spatial 기반 GNN(Graph Neural Network)으로, 기존 그래프 컨볼루션 신경망에서 Masked self multi-head 어텐션을 적용하여 제안된 신경망이다. 이는 다른 GNN과는 다르게 노드 분류를 수행하기 위한 어텐션 기반의 구조를 가지며, 인접한 노드의 정보만으로도 특징을 구할 수 있다. 즉, 완전히 미확인된 그래프를 테스트 셋으로 활용하는 것도 가능한 것이다. 본 연구에서는 이러한 특징을 가진 GAT를 통해 모든 센서를 각각 하나의 노드로 상정하여 센서 간 상관관계를 파악하고, 개별적 특징을 적용하여 엣지의 가중치를 다르게 적용하며 행동을 분류한다.

## III. The Proposed Scheme

본 논문에서 제안하는 반려동물 이상행동 탐지 서비스는 센서 데이터와 영상 데이터를 함께 활용하여 서로의 단점을 보완하고자 한다. 즉, 단일 데이터가 아닌 멀티모달

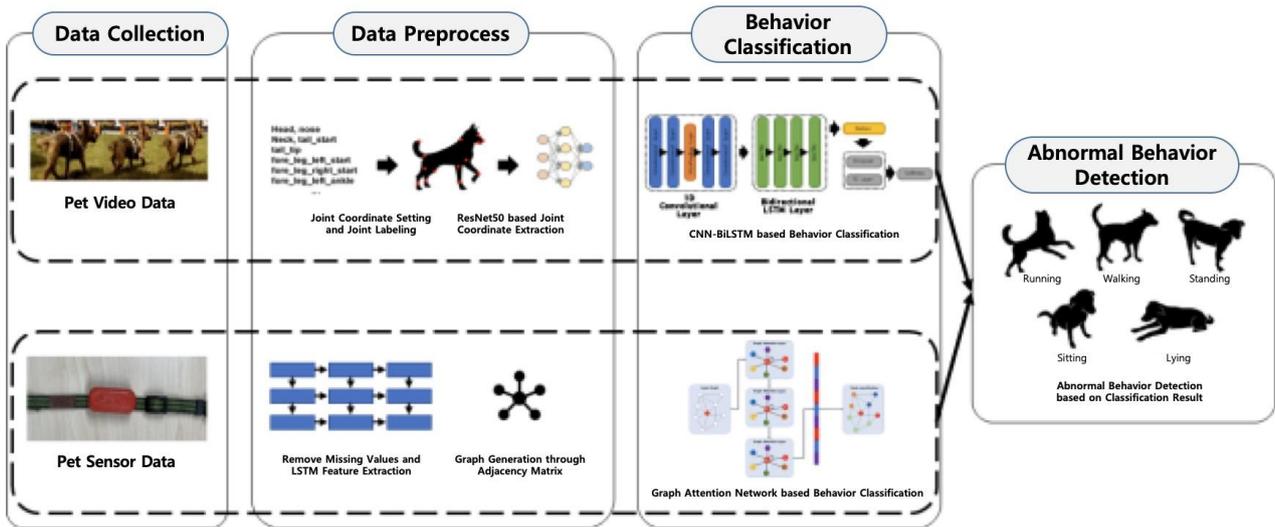


Fig. 1. System Architecture

데이터를 활용하는 것으로 행동 분류의 정확도를 개선하고, 이상행동을 보다 정확하게 탐지하는 것이 목적이다. 센서 데이터는 GAT를 통해 행동을 분류하고, 영상 데이터는 YOLO와 DeepLabCut을 활용하여 행동을 분류한다.

위 Fig. 1.은 멀티모달 데이터를 바탕으로 반려동물의 이상행동을 탐지하기 위한 모델의 개요이다. 본 연구에서 수집되는 데이터는 영상 데이터와 센서 데이터이며, 이를 각각 적합한 방식으로 처리하기 위해 서로 다른 데이터 전처리 과정을 거친다. 이후, 같은 행동에 대한 행동 분류를 바탕으로 이상행동을 탐지한다.

본 연구에서는 반려동물의 정상적 행동을 “걷다”, “뛰다”, “눕다”, “앉다”, “서다”의 5가지 행동으로 정의하였으며, 이외의 행동을 이상행동으로 분류하였다. 아래 그림 2.는 행동의 정의를 위해 번호를 부여한 것이다.



Fig. 2. Numbering for Classification of Behaviors

Fig. 2.에 부여한 신체 부위 번호를 바탕으로 Table 1.을 정의하였다.

Table 1. Behavior Classification Table

Action Group	Class	Number
Dynamic Behavior	“Running”	(1+2)-(3+4)
	“Walking”	1-3-2-4
Static Behavior	“Lying”	6 or (1+2+3+4+6)
	“Standing”	1+2+3+4
	“Sitting”	1+2+3+4+5

위 Table 1.의 기호 “+”는 해당하는 신체 부위가 동시에 땅에 닿는다는 의미이며, 기호 “-”는 해당 번호 순서대로 땅에 닿는다는 의미이다. 이때, 앞 차례 번호의 신체 부위는 땅에서 떨어진 상태이다.

### 3.1 Design of image data based behavior classification

영상 데이터를 통한 동물의 행동 분류를 위해 먼저 관절 좌표의 레이블링을 진행하였다. 관절 좌표의 레이블링 작업 이후 ResNet50을 활용하여 좌표의 학습을 진행하였다. ResNet50의 학습은 총 20,000번 진행되었으며 이를 통해 도출된 관절 좌표는 아래 Fig. 3.과 같다.

Joint Frame X, Y Estimated Coordinates	Joint Area						Likelihood
	bodyparts	head	head	head	nose	nose	
	x	y	likelihood	x	y	likelihood	
0	558.5621	380.3825	0.986705	538.0031	424.1452	0.965598	
1	558.5621	387.3757	0.804178	539.4774	426.092	0.59091	
2	558.5621	388.0366	0.98991	539.4774	426.8427	0.952493	
3	558.3977	390.8317	0.992455	539.6903	433.9892	0.989399	
4	558.0688	392.198	0.584328	539.7809	439.0773	0.796456	
5	558.0688	397.4969	0.173942	539.7809	443.8538	0.71638	
6	558.0688	397.8413	0.316213	539.8763	450.6749	0.555865	
7	558.0688	400.972	0.795918	539.8763	451.3276	0.833546	
8	561.6901	403.7688	0.907812	540.4929	456.6857	0.963183	
9	561.6901	403.7688	0.470836	540.4929	456.7813	0.866142	
10	556.6761	406.4637	0.934179	545.1643	457.2351	0.853336	

Fig. 3. Estimated Joint Coordinates

위 Fig. 3.의 가능도는 영상 프레임마다 설정된 신체에 레이블링한 좌표 값을 토대로 추론된 값이다. 본 연구에서는 데이터의 사용 여부를 50%의 상한선을 기준으로 설정하였으며, 가능도가 50% 이상으로 추론된 좌표들만 사용하였다.

위와 같은 과정으로 도출된 좌표들은 행동 분류를 위한 모델의 입력 값으로 사용된다. 입력 데이터는 총 13개의 관절에 해당하는 좌표이므로 총 26개의 칼럼으로 구성되어 있다. 특징을 추출하기 위해 출력 층이 256개인 1D Convolution의 2개 층을 지나서 한 번의 MaxPooling 과정을 거친 후, 다시 출력 층이 128개인 1D Convolution 계층을 지난다. 이후, 시계열 적 특징을 갖는 각 행동의 연관성을 파악하기 위한 BiLSTM 계층을 지난다. BiLSTM은 총 4개의 계층으로 구성되어 있다. 마지막 행동 분류를 위한 결과를 추출하기 위해 Flatten 계층 통해 전체 데이터를 1차원 벡터 형태로 구성하고, 다섯 가지의 행동을 분류하기 위해 Dense의 출력 층을 5로 설정한다.

### 3.2 Design of sensor data based behavior classification

센서 데이터 기반의 행동 분류는 영상 데이터와는 다른 데이터 전처리 과정을 거친다. 먼저, 시계열적 특성을 반영하기 위해 3초의 시간을 의미하는 3-layer LSTM 계층을 거쳐서 각 센서를 노드화하는 과정을 거친다. 아래 Fig. 4.는 이를 나타내는 그림이다.

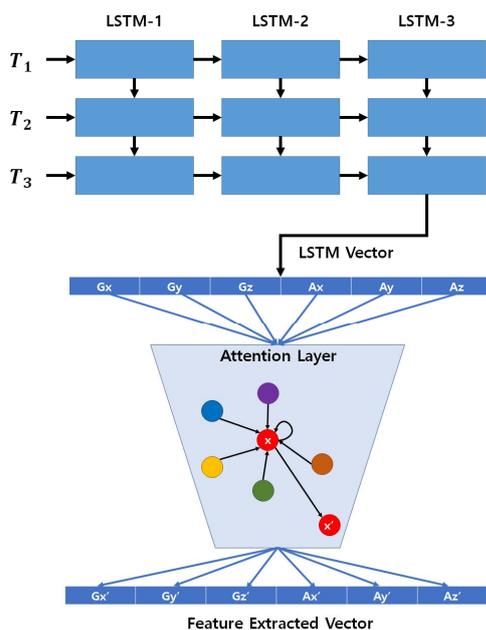


Fig. 4. LSTM Layer and Nodeization Phase

위와 같은 과정을 거쳐 각 센서 간의 상관관계를 파악하여 다시 그래프의 노드를 생성한다. 이후, 각 시간 별 노드의 인접 행렬을 생성하여 행동 분류를 위한 그래프를 생성한다. GAT 기반 행동 분류는 각 행동 간의 어텐션 계수를 바탕으로 노드를 분류한다. 본 연구에서는 3계층의 그래프 어텐션을 활용하였다. 이는 이웃한 노드와의 어텐션 계수를 바탕으로 새로운 노드를 생성한다. 이러한 과정의 그래프 임베딩 벡터에 대한 출력을 계산하여 분류가 이루어진다. 즉, 본 연구에서 활용되는 센서인 자이로 3축, 가속도 3축 각각의 상관관계와 각 행동의 상관관계를 고려하여 행동을 분류한다.

### 3.3 Abnormal behavior detection

센서 데이터와 영상 데이터를 바탕으로 행동을 분류한 후, 동물의 이상행동을 탐지한다. 분류되는 행동은 총 5가지로, “걷다”, “뛰다”, “앉다”, “서다”, “눕다”이며, 이는 2가지 대분류인 정적인 행동과 동적인 행동으로 나뉘볼 수 있다. 정적인 행동으로는 “앉다”, “서다”, “눕다”가 있으며, 동적인 행동으로는 “걷다”, “뛰다”가 있다. 이는 앞서 설명한 행동 정의 표에서 확인할 수 있다.

본 연구에서는 이러한 정적 행동 및 동적 행동을 기반으로 이상행동을 탐지한다. 같은 시간의 영상 분류 결과가 정적 행동으로 분류되었지만, 센서 분류 결과가 동적인 행동으로 분류되는 경우는 명확한 이상행동으로 분류한다. 하지만, 같은 동적인 행동 중 “걷다” 및 “뛰다”로 분류 결과가 다르게 나오는 경우, 영상 데이터를 우선으로 고려하여 해당하는 행동을 분류한다. 또한, 반려동물이 숨을 쉬는 것과 같이 발생하는 지속적이고 미약한 움직임은 노이즈로 상정하고 실험을 진행하였다. 마지막으로, 센서 데이터가 정적인 행동으로 분류하였으나 영상 데이터가 동적인 행동으로 분류할 경우 이는 이상행동이 아닌 센서의 결함으로 고려하여 추후 서비스 시에 확인할 수 있도록 신호를 전달한다. 아래 Table 2.는 이러한 내용을 정리한 내용이다.

Table 2. Abnormal Behavior Definition Table

Video Result	Sensor Result	Interpretation of Results
Static	Dynamic	Abnormal
Static	Static	Prioritize Video Result
Dynamic	Dynamic	
Dynamic	Static	Sensor Error Notification

## IV. Experiment and result

본 논문에서의 이상행동 분류는 각 분류 방법의 정확도가 중요하므로, 각 행동 분류 기법의 정확도를 파악하였다. 실험 환경은 Table 3.에 나타난 바와 같다.

Table 3. Experiment Environment

Type	Item
CCTV	Smart Home Camera Pro
Wearable Device	Arduino Nano 33 IoT Board
CPU	Intel i9-10980HK 2.4GHz
GPU	NVIDIA GeForce RTX 3080
RAM	32G
Python	3.6.8
Pytorch	1.7.1+cu110

### 4.1 Behavior classification based on image data

영상 데이터를 통한 행동 분류는 총 6,000개의 관절 좌표 프레임을 학습, 검증 및 테스트 데이터를 6:2:2의 비율로 구분하였다. 구분된 데이터는 입력 데이터로 설정하고, 분류 모델인 CNN-BiLSTM 모델과 아래 학습 파라미터를 적용하여 분류하였다.

Table 4. Image Classification Training Parameters

Parameter	Value
Learning rate	0.0005
Batch size	10
Output dimension	6
Epoch	100 (Early stopping = 10)
Loss function	Categorical crossentropy
Optimizer	Adam

위 학습 파라미터를 적용하여 행동을 분류하였으며, 영상 분류의 경우 분류성능평가지표인 정밀도, 재현율, 그리고 F1-score를 통해 성능을 평가하였다. 아래 Fig. 5.와 Fig. 6.은 각각 모델의 학습 그래프를, Table 5.는 행동별 분류 결과를 나타낸다.

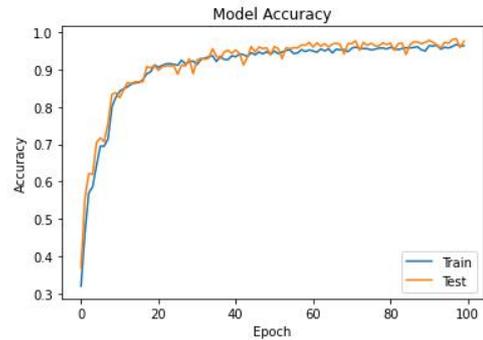


Fig. 5. Image Classification Model Training Graph (Accuracy)

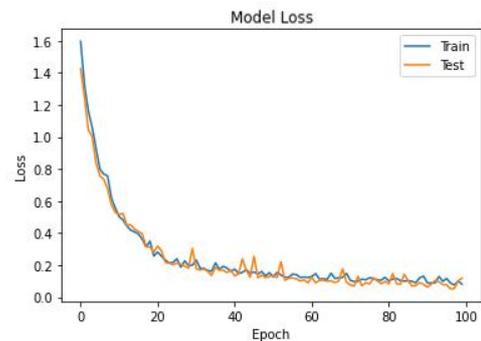


Fig. 6. Image Classification Model Training Graph (Loss)

Table 5. Behavior Classification Result

Classified Behavior	Accuracy	Recall	F1-score
Running	0.94	0.85	0.89
Walking	0.94	0.88	0.91
Standing	0.99	0.96	0.98
Sitting	1.00	1.00	1.00
Lying	1.00	1.00	1.00

### 4.2 Behavior classification based on sensor data

센서 데이터를 바탕으로 이루어지는 행동 분류는 총 3,324초의 데이터를 학습하여 진행되었다. 이는 영상을 확인하며 시간마다 행동을 라벨링한 데이터이다. 그 중, 결측치는 직전 및 직후 센서 측정값의 평균치로 처리하였다. 또한, 전처리의 과정 중 그래프 생성 과정에서 아무런 노드와도 연결되지 않으며 동떨어지는 노드는 이상치로 고려하여 제외하였다.

그래프를 생성하는 과정에서 엣지의 생성은 피어슨 상관 계수를 활용하였다. 피어슨 상관계수는 값이 서로 함께 변하는 정도를 파악하며, 본 연구에서는 이러한 수치가 0.5 이상을 보일 때 엣지를 연결하여 그래프를 생성하였다.

본 논문에서는 3개의 어텐션 계층을 쌓았으며, 입력 값

으로 활용된 그래프는 672,624개의 엣지와 2,173개의 노드를 가지고 있다. 아래 Fig. 7.은 구현 모델의 요약이며, Table 6.은 적용된 학습 파라미터이다.

Layer (type)	Output Shape	Param #
dense_30 (Dense)	multiple	14000
multi_head_graph_attention_4	multiple	4004000
multi_head_graph_attention_4	multiple	4004000
multi_head_graph_attention_4	multiple	4004000
dense_31 (Dense)	multiple	10005

Total params: 12,036,005  
 Trainable params: 12,036,005  
 Non-trainable params: 0

Fig. 7. Model Summary

Table 6. Graph Classification Parameter

Parameter	Value
Hidden Units	1000 (About half of the nodes)
Learning Rate	0.00001
Batch Size	2
Output Dimension	5
Epoch	100 (Early stopping = 10)
Loss Function	Sparse Categorical Crossentropy
Optimizer	SGD (Stochastic Gradient Descent)

테스트 데이터 셋은 총 200개의 데이터 셋을 활용하였으며, 아래 Fig. 8.과 Fig. 9.는 학습 그래프를, Fig. 10.은 분류 결과를 나타낸다.

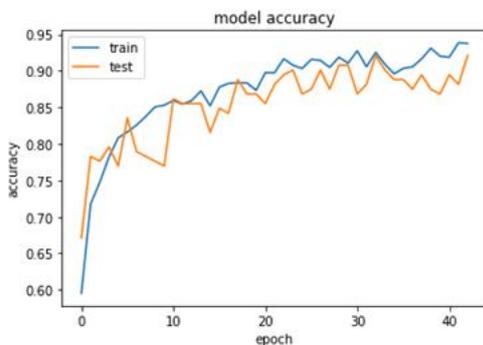


Fig. 8. Sensor-based Classification Model Training Graph(Accuracy)

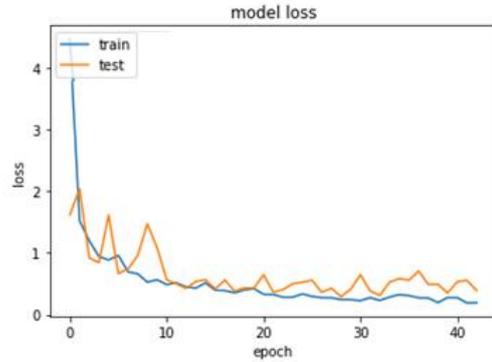


Fig. 9. Sensor-based Classification Model Training Graph(Loss)

```

Test data Answer : standing
Probability of lyingDown behavior = 10.415%
Probability of running behavior = 0.000%
Probability of sitting behavior = 7.412%
Probability of standing behavior = 82.152%
Probability of walking behavior = 0.021%
-----
Test data Answer : walking
Probability of lyingDown behavior = 0.000%
Probability of running behavior = 11.275%
Probability of sitting behavior = 1.537%
Probability of standing behavior = 0.047%
Probability of walking behavior = 87.141%
-----
Test data Answer : sitting
Probability of lyingDown behavior = 0.015%
Probability of running behavior = 0.000%
Probability of sitting behavior = 93.579%
Probability of standing behavior = 6.192%
Probability of walking behavior = 0.214%
-----
Test data Answer : lyingDown
Probability of lyingDown behavior = 84.594%
Probability of running behavior = 0.021%
Probability of sitting behavior = 15.065%
Probability of standing behavior = 0.118%
Probability of walking behavior = 0.202%
-----
Test data Answer : running
Probability of lyingDown behavior = 0.000%
Probability of running behavior = 99.926%
Probability of sitting behavior = 0.000%
Probability of standing behavior = 0.000%
Probability of walking behavior = 0.074%
-----
Test Accuracy 91.1%
    
```

Fig. 10. Classification Results and Accuracy

## V. Conclusions

본 논문에서는 멀티모달 데이터를 활용한 반려동물의 이상행동 탐지 서비스를 설계 및 구현하였다. 주변에서 흔히 볼 수 있는 CCTV를 통한 영상 데이터와 목줄형 웨어러블 디바이스를 통해 센서 데이터를 수집하였다. 영상 데이터는 객체 추적 및 추정 좌표 추출을 바탕으로 높은 정확도로 반려동물의 행동을 분류하였으며, 센서 데이터는 GAT를 통해 각 센서의 상관관계를 반영하여 영상과 마찬가지로 높은 정확도로 행동을 분류하였다.

하지만, 본 연구는 충분한 이상행동 데이터를 수집하지 못했기 때문에 더욱 세밀한 이상행동을 탐지할 수 없다는 한계가 있다. 또한, 센서 데이터를 통한 분류의 경우, “높

다”를 포함한 “걷다”와 “서다”의 분류 정확도가 낮은 것을 확인할 수 있었다. 이와 같은 결과는 데이터의 부족과 센서 데이터를 통한 큰 동작이 아닌 작은 동작의 분류가 더욱 정확하게 이루어지지 않았기 때문이다. 이는 추후 전문가의 의견이 반영된 이상행동 데이터를 수집하는 것과 센서 데이터와 영상 데이터의 결합 과정으로 더욱 세밀하고 정확한 이상행동을 탐지할 수 있을 것으로 기대된다.

추후 추가로 영상과 센서뿐만이 아닌 소리 또한 추가적인 정보로 활용하여 더욱 정확한 행동 분류 및 이상행동 탐지를 위한 정보로 활용할 계획이다.

## ACKNOWLEDGEMENT

This research was supported by the Academic Research Fund of Hoseo University in 2020 (20200814)

## REFERENCES

- [1] S. Kang, H. Kim, Y. Ji and J. Lee, “A Study on the Acceptance Factors of the Introduction of a Smart IoT Technology for Well-being Companion Animal,” *The Journal of Society for e-Business Studies*, 24, 2, pp. 143-163, March 2019. DOI:10.7838/jsebs.2019.24.2.143
- [2] J. Lee and D. Kim, “Implementation of a Smart IoT System with Automatic Pet Feeder,” *Journal of Digital Contents Society*, 22, 2, pp. 209-214, February 2021. DOI : 10.9728/dcs.2021.22.2.209
- [3] W. Kim, S. Kim, and H. Kwak, “The Companion Animal Monitoring System using Low-Power Protocol Wearable Device,” *Journal of the Korea Society of Computer and Information*, 25, 12, pp.17-23, December 2020. DOI : 10.9708/jksoci.2020.25.12.017
- [4] H. Shin, J. Kim, “A Study on the Development of Wearable Products Applied to PetTech Service Using IoT and AI Technology,” *JOURNAL OF THE KOREAN SOCIETY DESIGN CULTURE*, 26, 1, pp. 261-272, March 2020. DOI : 10.18208/ksdc.2020.26.1.261
- [5] K. Pekka, V. Anna, S. Sanni, T. Heini, V. Heli, M. Päivi, G. Yulia, A. Christoph, S. Veikko, K. Miiamaaria, V. outi, and V. Antti, “Dog behaviour classification with movement sensors placed on the harness and the collar,” *Applied Animal Behaviour Science*, 241, pp.1-7, June 2021. DOI: 10.1016/j.applanim.2021.105393
- [6] H. Kim, N. Choi, J. Kang, S. Lim and H. Kim, “Real-time Face Recognition System in CCTV Video for Crime Prevention,” *The Journal of Korean Institute of Information Technology*, 19, 8, pp.99-106, August 2021. DOI:10.14801/jkiit.2021.19.8.99
- [7] M. Kröschel, B. Reineking, F. Werwie, F. Wildi and I. Storch. “Remote monitoring of vigilance behavior in large herbivores using acceleration data,” *Animal Biotelemetry*, Vol.5, No.1, pp.1-15, May 2017. DOI: 10.1186/s40317-017-0125-z
- [8] L. Yang, D. Jiang, X. Xia, E. Pei, C. Oveneke and H. Sahli, “Multimodal measurement of depression using deep learning models,” In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, pp. 53-59, October 2017. DOI: 10.1145/3133944.3133948
- [9] J. Joshi, R. Goecke, S. Alghowinem, A. Dhall, M. Wagner, J. Epps, P. Gordon and M. Breakspear, “Multimodal assistive technologies for depression diagnosis and monitoring,” *Journal on Multimodal User Interfaces*, 7, 3, pp.217-228, September 2013. DOI: 10.1007/s12193-013-0123-2
- [10] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, “You only look once: Unified, real-time object detection,” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.779-788, June 2016. DOI: 10.1109/CVPR.2016.91
- [11] D. Wu, S. Lv, M. Jiang and H. Song, “Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments,” *Computers and Electronics in Agriculture*, 178, pp.1-12, August 2020. DOI:10.1016/j.compag.2020.105742
- [12] J. Yu, and Z. Wei, “Face mask wearing detection algorithm based on improved YOLO-v4,” *Sensors*, 21, 9, pp.1-21, May 2021. DOI: 10.3390/s21093263
- [13] P. Mahto, P. Garg, P. Seth and J. Panda, “Refining Yolov4 for vehicle detection,” *International Journal of Advanced Research in Engineering and Technology*, 11, 5, pp.409-419, July 2020. DOI:10.34218/IJARET.11.5.2020.043
- [14] A. Bewley, G. Zongyuan, O. Lionel, R. Fabio and U. Ben, “Simple online and realtime tracking,” In *2016 IEEE international conference on image processing*, pp.3464-3468, July 2016. DOI: 10.48550/arXiv.1602.00763
- [15] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka and B. Schiele, “Deepcut: A deeper, stronger, and faster multi-person pose estimation model,” *European Conference on Computer Vision*, pp.34-50, September 2016. DOI:10.1007/978-3-319-46466-6\_4\_3
- [16] H. Xu, L. Shen, Q. Zhang and G. Cao, “Fall behavior recognition based on deep learning and image processing,” *International Journal of Mobile Computing and Multimedia Communications*, 9, 4, pp.1-15, October 2018. DOI: 10.4018/IJMCMC.2018100101
- [17] W. Chenhao, W. Yongquan, G. Dong and G. Jun, “Human behavior recognition under occlusion based on two-stream

network combined with BiLSTM,” In 2020 Chinese Control And Decision Conference (CCDC), pp. 3311-3316, August 2020. DOI: 10.1109/CCDC49329.2020.9164760

- [18] R. Vincent, A. Wagadre, A. Sivaraman, M. Rajesh and A. Rajesh, “Human Activity Recognition Using LSTM/BiLSTM,” *International Journal of Advanced Science and Technology*, 29, 4, pp.7468-7474. April 2020.
- [19] R. Arablouei, L. Wang, L. Currie, F. Alvarenga, and G. Bishop-Hurley, “Animal Behavior Classification via Deep Learning on Embedded Systems,” *arXiv preprint arXiv:2111.12295*, pp.1-13, November 2021. DOI: 10.48550/arXiv.2111.12295
- [20] K. Sakai, K. Oishi, M. Miwa, H. Kumagai and H. Hirooka, “Behavior classification of goats using 9-axis multi sensors: The effect of imbalanced datasets on classification performance,” *Computers and Electronics in Agriculture*, 166, pp.1-9, September 2019. DOI:10.1016/j.compag.2019.105027
- [21] N. Kleanthous, A. Hussain, W. Khan, J. Sneddon and A. Mason, “Feature Extraction and Random Forest to Identify Sheep Behavior from Accelerometer Data” *International Conference on Intelligent Computing*. Springer, pp.408-419, October 2020. DOI:10.1007/978-3-030-60796-8\_35
- [22] M. Hassoon, M. Kouhi, M. Zomorodi-Moghadam and M. Abdar, “Rule optimization of boosted c5. 0 classification using genetic algorithm for liver disease prediction,” In 2017 international conference on computer and applications, IEEE, pp.299-305, October 2017. DOI: 10.1109/COMAPP.2017.8079783
- [23] M. Salman Saeed, M. Mustafa, U. Sheikh, T. Jumani, I. Khan, S. Atawneh and N. Hamadneh, “An efficient boosted C5. 0 Decision-Tree-Based classification approach for detecting non-technical losses in power utilities,” *Energies*, 13, 12, pp.1-19, June 2020. DOI:10.3390/en13123242

## Authors



Ji-Hoon Lee received the B.S. and M.S. degrees in Computer Science and Engineering from Hoseo University, Korea, in 2020, 2022. He is interested in deep learning, big data processing and analysis, and graph neural networks.



Min-Chan Shin received the B.S. and M.S. degrees in Computer Science and Engineering from Hoseo University, Korea, in 2020, 2022. He is interested in deep learning, big data processing and analysis, image processing, and object estimation.



Jun-Hee Park received the B.S. and M.S. degrees in Computer Science and Engineering from Hoseo University, Korea, in 2020, 2022. He is interested in deep learning, big data processing and analysis, and multi-modal analysis.



Nam-Mee Moon received B.S., M.S., and Ph.D. degrees in School of Computer Science and Engineering from Ewha Womans University in 1985, 1987 and 1998, respectively.

She served as an assistant professor at Ewha Womans University from 1999 to 2003. From 2003 to 2008, she is a professor of Department Digital Media, Graduate School of Seoul Venture Information. Since 2008, she is currently a professor in the Department of Computer Science and Engineering, Hoseo University. She is current research interests include Social Learning, HCI and User Centric Data, Big-data Processing and Analysis.