JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 3, pp. 107-116, March 2026
https://doi.org/10.9708/jksci.2026.31.03.107

Edge-Cloud Pet Behavior Monitoring System Based on On-Device Encoding Anonymization

Hyuksoon Choi*, JinHwan Yang*, Nammee Moon**

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

[Abstract]

This paper proposes an adversarial learning-based privacy-preserving behavior monitoring system that fundamentally eliminates sensitive information at the edge device level. The proposed system adopts a Masked Autoencoder (MAE) as a backbone to effectively learn contextual features of behavior even in environments with limited labeled data. Specifically, we introduce a confusion loss into the encoding process to perform Min-Max optimization, which preserves the utility information required for behavior analysis while minimizing identity-specific information. Experimental results demonstrate that the proposed model maintains superior behavior classification accuracy compared to existing baseline models while significantly reducing identity re-identification F1 Score to a random guess level of 0.087(Behavior Classification), 0.052(Audio Detection) thereby validating the security and practicality of the system.

▸ **Key words**: Pet Monitoring System, Privacy Protection, Self-Supervised Learning, Data Anonymization

[요 약]

본 논문은 엣지 디바이스 단계에서 민감 정보를 원천적으로 소거하는 적대적 학습 기반의 프라이버시 보존형 행동 모니터링 시스템을 제안한다. 제안하는 시스템은 마스킹된 오토인코더(MAE)를 백본으로 채택하여 레이블이 부족한 데이터 환경에서도 행동의 문맥적 특징을 효과적으로 학습한다. 특히, 인코딩 과정에 프라이버시 손실을 도입하여, 행동 분석에 필요한 유용성(Utility) 정보는 보존하되 개체 고유의 식별 정보는 최소화하는 Min-Max 최적화를 수행한다. 실험 결과, 제안 모델은 기존 베이스라인 모델 대비 우수한 행동 분류 정확도를 유지하면서도, 개체 재식별 F1 점수를 무작위 추측 수준인 0.087(행동 분류), 0.052(음성 탐지)까지 낮추어 시스템의 보안성과 실용성을 입증하였다.

▸ **주제어**: 반려동물 모니터링 시스템, 프라이버시 보호, 자기지도 학습, 데이터 익명화

• First Author: Hyuksoon Choi, Corresponding Author: Nammee Moon
*Hyuksoon Choi (hucksoon2001@gmail.com), Dept. of Computer Science and Engineering, Hoseo University
*JinHwan Yang (yjh970706@naver.com), Dept. of Computer Science and Engineering, Hoseo University
**Nammee Moon (nammee.moon@gmail.com), Dept. of Computer Science and Engineering, Hoseo University
• Received: 2026. 01. 19, Revised: 2026. 02. 24, Accepted: 2026. 03. 05.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

최근 1인 가구의 증가와 펫코노미(Petconomy) 시장의 성장으로 인해, 가정 내 반려동물의 상태를 실시간으로 확인하는 지능형 모니터링 시스템의 수요가 급증하고 있다[1]. 기존의 반려동물 모니터링 서비스는 주로 가정 내 설치된 IP 카메라나 IoT 센서가 수집한 영상 및 IMU 센서 데이터를 클라우드 서버로 전송하여, 반려동물의 행동 패턴을 딥러닝 모델로 분석하는 중앙 집중형 방식을 따른다[2]. 이러한 접근 방식은 고성능 연산 자원을 활용하여 높은 분석 정확도를 제공한다는 장점이 있으나, 가정 내부의 민감한 원본 데이터를 외부 서버로 전송해야 한다는 점에서 잠재적인 보안 위협을 내포하고 있다.

최근 연구들에 따르면, 원본 데이터가 아닌 딥러닝 모델의 중간 출력값인 잠재 특징 벡터만을 전송하더라도, 이를 역추적하여 원본 정보를 복원하거나 해당 데이터가 특정 개체의 것임을 식별하는 재식별 공격이 가능함이 입증되었다[3]. 특히 반려동물 모니터링 환경에서 행동 정보뿐만 아니라 반려동물 음성 데이터 등에 포함된 개체 식별 정보가 서버로 유출될 경우, 이는 단순한 데이터 유출을 넘어 사용자 및 반려견의 신원 노출로 이어질 수 있다[4,5]. 따라서 행동 인식을 위한 정보는 보존하면서도, 개체를 식별할 수 있는 민감 정보는 선택적으로 제거하는 온디바이스 익명화 기술이 필수적이다[6].

이에 본 논문에서는 프라이버시 손실 기반의 적대적 학습 기법을 적용하여, 엣지 디바이스(Edge Device) 단계에서 개체 식별 정보를 효과적으로 소거하는 엣지-클라우드 기반 프라이버시 보존형 반려동물 행동 모니터링 시스템을 제안한다. 제안하는 인코더는 행동 분류기의 성능을 유지하는 동시에, 프라이버시 분류기의 예측 확률이 모든 개체에 대해 균등 분포를 따르도록 학습된다. 이를 통해 서버로 전송되는 잠재 특징 벡터는 무작위 추측 수준의 익명성을 확보하게 되며, 결과적으로 공격자가 전송 구간에서 데이터를 탈취하여 역공학을 시도하더라도 식별 가능한 원본 정보를 복원할 수 없도록 한다.

본 논문의 구성은 다음과 같다. 2장에서 프라이버시 위협과 방어 기법과 프라이버시 보존형 모니터링 시스템에 관련된 연구를 설명하고 3장에서 제안하는 반려동물 모니터링 시스템과 대해 설명한다. 4장에서 앞서 제안한 시스템을 바탕으로 실험을 진행하며, 마지막으로 5장에서 결론 및 향후 연구를 제시한다.

II. Preliminaries

1. Related works

심층 신경망(Deep Neural Networks)은 데이터의 고수준 특징을 추출하는 데 탁월한 성능을 보이지만, 이 과정에서 학습 데이터의 민감한 개체 식별 정보가 잠재 특징 벡터에 내재화되는 문제가 지속적으로 제기되어 왔다. 초기 연구인 A. Dosovitskiy et al.은 특징 역전파를 통해 시각적 정보를 복원할 수 있음을 증명하였으며, 이는 딥러닝 보안의 기초적인 위협 모델로 자리 잡았다[7].

최근 연구들은 이러한 위협이 더욱 정교해지고 있음을 경고하고 있다. N. Carlini et al.은 확산 모델(Diffusion Model)과 같은 생성형 AI 기술을 활용하여 극소량의 특징 정보만으로도 원본 데이터를 고해상도로 복원하는 생성형 역전파 공격을 시연하였다[8]. 또한, J. Zhang et al.은 엣지-클라우드(Edge-Cloud) 협업 환경에서 중간 특징값의 통계적 특성만을 분석하여 사용자의 신원을 추론하는 속성 추론 공격의 위험성을 입증하였다[9]. 더 나아가, 분할 학습(Split Learning) 환경에서도 원본 입력 대신 중간 활성화만 전송하더라도, 서버가 이를 이용해 입력을 복원하거나 민감 속성을 추론할 수 있다는 점이 체계적으로 보고되고 있다[10].

특히, 최근에는 공격자가 아주 제한된 양의 보조 데이터만 보유하고 있더라도, 피해 클라이언트의 특징 표현 경향을 역으로 학습하여 중간 특징값으로부터 원본 데이터를 고품질로 재구성해내는 FORA와 같은 고도화된 공격 기법들이 보고되고 있다[11].

이러한 위협에 대한 방어책으로 차분 프라이버시가 오랫동안 표준으로 사용되었으나, 이는 데이터의 효용성이 급격히 저하되는 한계가 있다. 이를 보완하기 위해 M. Gong et al.은 특징 채널별로 노이즈를 적응적으로 할당하는 메커니즘을 제안하였으나, 여전히 정밀한 작업에서는 성능 저하가 발생함이 보고되었다[12]. 이에 따라 최근에는 적대적 학습 기반의 방어가 주목받고 있다. P. Vepakomma et al.은 정보 병목 이론과 적대적 학습을 결합하여, 민감 정보는 억제하고 타겟 정보는 최대화하는 최적화 프레임워크를 제안하였다[13].

이에 본 논문은 이러한 적대적 방어 기법을 발전시켜, 프라이버시 손실을 통해 개체 식별 확률을 무작위 추측 수준으로 유도함으로써, 공격자가 데이터를 탈취하더라도 개체 정보를 복원할 수 없도록 하는 시스템을 제안한다.

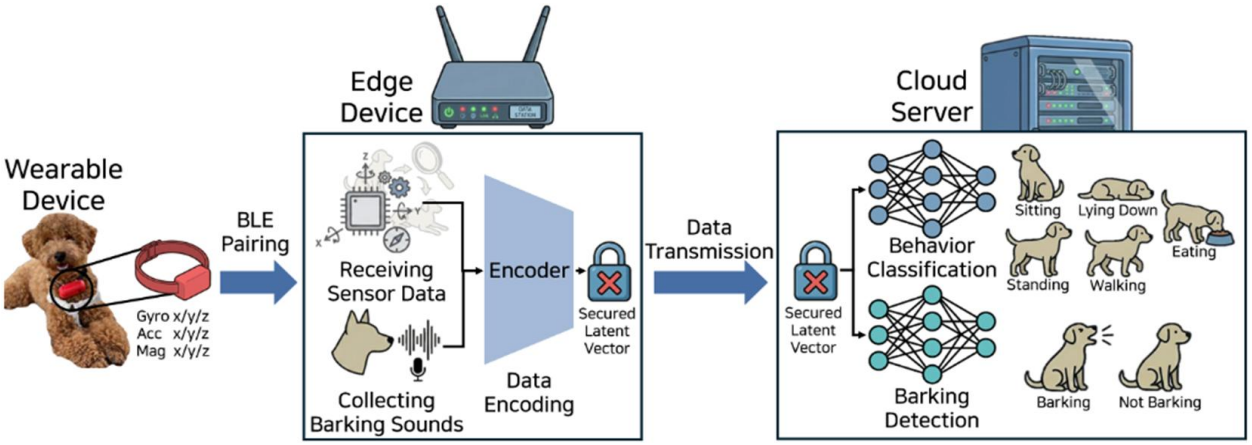

Fig. 1. Overview of edge-cloud pet behavior monitoring system based on on-device anonymization

2. Privacy-protecting monitoring system

IoT 센서 데이터의 폭발적인 증가로 인해, 데이터를 클라우드 서버로 전송하지 않고 엣지 디바이스와 협업하는 분할 컴퓨팅 기술이 발전하고 있다. Y. Kang et al.의 Neurosurgeon이 제안한 초기 분할 컴퓨팅 개념은 연산 효율성에 초점을 맞추었으나, 최근 연구들은 이를 프라이버시 보호 도구로 재해석하고 있다[14].

초기에는 단순히 신경망의 레이어를 나누는 방식이 주를 이뤘으나, 최근 D. Yao et al.은 단순한 분할 학습 구조에서도 역전파를 통해 원본 데이터가 재구성될 수 있음을 지적하며, 전송되는 잠재 특징 벡터 자체에 대한 보안 메커니즘이 필수적임을 강조하였다[15]. 대표적으로 M. Caprolu et al.은 SCIPER 프레임워크를 통해, 엣지 디바이스의 특징 추출 과정에 별도의 정규화 손실 함수를 도입하여 원본 데이터와의 유사성을 최소화하면서도 타겟 작업의 성능을 유지하는 방법을 제안하였다[16]. 또한, H. Xu et al.는 Privacy-Preserving Convolutional Autoencoder 모델을 통해 엣지-클라우드 간 데이터 전송 시 실제 행동 정보는 유지하되 환경 노이즈와 민감한 배경 정보를 억제하는 오토인코더 기반의 전송 구조를 설계하였다[17].

반려동물 모니터링 분야에서도 P. Kumulainen et al.과 R. D. Chambers et al.이 웨어러블 센서를 활용한 정밀 행동 인식 모델을 제안하였으나, 대부분은 데이터 보안보다는 모델의 경량화나 인식 정확도 향상에 초점을 맞추고 있다[18,19]. 본 논문은 단순한 정규화나 오토인코딩을 넘어 적대적 학습 기반의 프라이버시 손실을 적용하여 개체 식별 확률을 무작위 수준으로 강제함으로써, 보안 강도와 유틸리티 보존 측면에서 더욱 효과적인 프레임워크를 제안한다.

III. The Proposed Method

1. Overview of proposal system

본 논문에서는 반려동물 모니터링 과정에서 발생하는 데이터 유출 문제를 해결하기 위해, 온디바이스 익명화 기반 엣지-클라우드 반려동물 행동 모니터링 시스템을 제안한다. 제안하는 시스템의 개요도는 Fig. 1.과 같다.

제안하는 시스템은 크게 데이터 수집 및 온디바이스 익명화를 담당하는 엣지 디바이스와, 수신된 데이터를 바탕으로 심층 분석을 수행하는 클라우드 서버(Cloud Server)로 계층화된다. 이러한 구조적 분리는 민감 정보를 포함한 원본 IMU 센서 데이터 및 반려동물 음성 데이터와, 서비스 제공을 위해 가공된 특징 데이터를 물리적으로 격리하는 역할을 한다. 이를 위해 엣지 디바이스는 원본 데이터를 외부로 전송하지 않고, 자체적으로 특징 추출 과정을 수행한 후, 정제된 잠재 특징 벡터만을 클라우드 서버로 전송한다.

특히, 본 시스템은 데이터의 유용성을 유지하면서 개체 식별 정보를 선택적으로 제거하기 위해, 적대적 학습 기반의 최적화 전략을 도입한다. 엣지 디바이스 내의 인코더는 행동 분류기의 손실을 최소화하여 분석 정확도를 유지하는 동시에, 프라이버시 분류기(Privacy Head)의 손실을 최대화하는 Min-Max 최적화 방향으로 학습된다.

결과적으로 클라우드 서버로 전송되는 데이터는 행동 패턴 분석에 필요한 정보량은 보존하되, 개체 고유의 식별 정보는 수학적으로 소거된 상태가 된다. 이는 공격자가 전송 구간에서 데이터를 탈취하더라도, 원본 신호로의 복원이나 개체에 대한 재식별 공격을 불가능하게 만들어 시스템 전반의 보안성을 획기적으로 향상시킨다.

2. Encoder for on-device anonymization

본 절에서는 사용자 가정 내에 설치된 엣지 디바이스에서 수행되는 데이터 수집 및 온디바이스 처리 과정을 상세히 기술한다.

엣지 디바이스는 반려동물로부터 수집된 고차원의 IMU 센서 데이터와 반려동물 음성 데이터를 효율적으로 처리하기 위해, 입력 데이터를 패치 단위로 분할하고 일정 비율로 마스킹한 후, 관측 가능한 패치만을 추출하여 인코딩하는 과정을 수행한다. 본 시스템은 연산 자원이 제한적인 온디바이스 환경을 고려하여, 입력 데이터의 중복성을 제거하고 핵심 정보만을 압축적으로 학습하는 경량화된 MAE(Masked Autoencoder) 기반의 인코더를 백본으로 채택하였다. 데이터 처리 과정은 모달리티에 따라 개별적인 모델을 통해 패치 분할, 무작위 마스킹, 그리고 특징 인코딩의 단계로 진행된다.

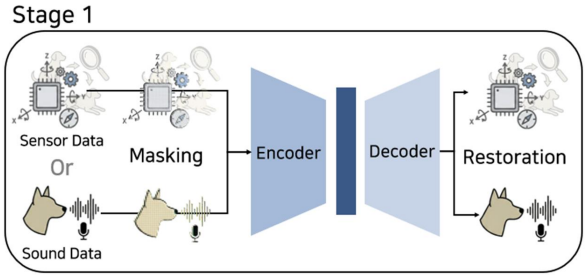

Fig. 2. Overview of self-guided pre-training process

첫째, 가속도, 자이로스코프, 지자계로 구성된 IMU 센서 데이터는 1차원 시퀀스 형태로 처리된다. 센서 모델은 연속적인 입력 신호를 고정된 길이의 패치로 분할하여 지역적인 움직임 패턴을 포착한다. 마스킹되지 않고 남은 패치들은 선형 투영 층을 거쳐 고차원의 임베딩 벡터로 변환되며, 이때 시계열 데이터의 순차적 정보를 보존하기 위해 학습 가능한 위치 임베딩이 더해진다. 이를 통해 센서 인코더는 부분적인 정보만으로도 반려동물의 미세한 움직임과 전역적 행동 흐름을 함축하는 잠재 특징을 효과적으로 추출한다.

둘째, 반려동물 음성 데이터는 별도의 오디오 처리 모델을 통해 분석된다. 원본 음성 파형을 시간과 주파수 정보를 동시에 포함하는 2차원 스펙트로그램(Spectrogram)으로 변환한 후 모델에 입력한다. 음성 모델은 스펙트로그램을 센서 처리 방식처럼 고정된 길이의 패치로 분할하여 지역적인 주파수 패턴을 포착한다. 이후 높은 비율의 무작위 마스킹을 통해 모델이 오디오 데이터의 주요 특성을 포함하는 임베딩 벡터를 추출할 수 있도록 학습한다. 이를 통해 오디오 처리 모델은 배경 소음 속에서도 반려동물의 짖는 소리를 효과적으로 포착한다.

이와 같은 자기지도 학습 과정을 통해, 각 모달리티의 인코더는 레이블이 없는 대규모 데이터셋으로부터 반려동물 행동과 음성에 내재된 고유한 분포 및 문맥적 연관성을 함축적으로 학습한다. 본 시스템은 이렇게 사전 학습된 인코더의 파라미터를 다운스트림 작업인 행동 분류 및 음성 탐지를 위한 초기 가중치로 활용한다. 자기지도 학습 과정의 개요도는 Fig. 2와 같다.

다음으로 미세조정(Fine-tuning) 단계에서는 구조적인 최적화가 수행된다. 미세조정 단계에는 인코더 출력은 동시에 유틸리티 헤드와 프라이버시 헤드로 전달된다. 유틸리티 헤드는 행동 또는 음성 이벤트를 예측하며, 프라이버시 헤드는 동일한 잠재 벡터로부터 개체 ID를 예측한다. 학습 시 유틸리티 헤드는 분류 오차를 최소화하도록, 프라이버시 헤드는 개체 식별 정확도를 높이도록 학습되지만, 인코더는 반대로 프라이버시 헤드가 올바른 ID를 예측하지 못하도록 업데이트된다. 따라서 적대적 학습은 인코더와 프라이버시 헤드 사이의 경쟁 구조에서 수행된다.

사전 학습 단계에서 마스킹된 데이터를 원본으로 복원하는 역할을 수행했던 디코더는 연산 효율성을 위해 제거되며, 오직 인코더만이 엣지 디바이스의 특징 추출 백본으로 채택된다. 이후 인코더는 데이터의 복원이 아닌, 행동 분석에 최적화된 고차원의 잠재 특징 벡터를 추출하도록 미세 조정된다.

구체적으로, 프라이버시 손실을 활용한 적대적 학습 프레임워크를 통해, 잠재 벡터로부터 행동을 분류하는 유틸리티 헤드(Utility Head)와 개체를 식별하는 프라이버시 헤드를 동시에 운용하며, 두 헤드 간의 경쟁적인 학습을 유도한다. 이때 인코더는 유틸리티 헤드의 분류 오차를 최소화하여 행동 인식 성능을 유지하는 반면, 프라이버시 헤드에 대해서는 예측 확률 분포가 모든 개체에 대해 균등한 확률을 갖는 균등 분포를 따르도록 강제함으로써 개체 식별 불확실성을 최대화한다. 이때 목적 함수는 수식 (1)과 같으며 행동 분류를 위한 유틸리티 손실 수식 (2)와 프라이버시 보호를 위한 프라이버시 손실 수식 (3)의 가중합으로 정의된다.

$$J(\theta) = \min_{\theta_e,\theta_u} \max_{\theta_p} L_{utility}(\theta_e,\theta_u) - \lambda L_{privacy}(\theta_e,\theta_p) \quad (1)$$

$$L_{uyility} = \frac{1}{N}\sum_{i=1}^{N} \ell_{cls}(y_i, z_i) \quad (2)$$

$$L_{proivacy} = \frac{1}{|B|}\sum_{i\in B} CE(p_i, sotfmax(q_i)) \quad (3)$$

수식 (1)은 θ_e, θ_u, θ_p가 각각 사전학습된 인코더, 유틸리티 헤드, 프라이버시 헤드의 파라미터일 때 제안 모델의 목적함수 $J(\theta)$이다. λ는 $L_{privacy}$의 적용 정도를 조절하는 하이퍼 파라미터이다. 해당 수식을 통해 제안 모델은 $L_{utility}$를 줄이면서 $L_{privacy}$를 증가시키는 방향으로 학습한다.

$L_{utility}$는 수식 (2)를 따른다. y_i는 행동에 대한 라벨을 뜻하고 z_i는 유틸리티 헤드의 예측값을 뜻한다. ℓ_{cls}는 분류 과제에 맞는 손실 함수를 뜻한다. 행동 분류는 교차 엔트로피를 사용하고 음성 탐지는 이진 교차 엔트로피를 사용한다. $L_{privacy}$는 수식 (3)을 따른다. B는 각 샘플 집합, p_i는 민감 정보 라벨, q_i는 프라이버시 헤드의 예측값을 뜻하며, 라벨과 예측값의 교차 엔트로피를 사용한다. $L_{privacy}$를 증가시키도록 인코더를 학습한다는 것은, 프라이버시 헤드가 잠재벡터로부터 개체 ID를 안정적으로 예측하지 못하도록 표현 공간을 재구성한다는 의미이다. 그 결과 잠재벡터에는 행동 분류에 필요한 공통 동작 패턴은 유지되지만, 개체별 보행 습관, 체형, 음성 톤과 같이 재식별에 기여하는 미세 특징은 억제되거나 교란된다. 즉, $L_{privacy}$의 증가는 단순한 손실값 증가가 아니라, 공격자 관점에서 잠재 표현의 식별 가능성을 낮추는 비식별화 효과로 해석될 수 있다.

이러한 Min-Max 최적화 과정을 통해, 엣지 디바이스는 최종적으로 행동 분석에 필요한 정보량은 보존하되 개체 고유의 생체적 특징은 소거된 익명화된 특징 벡터만을 생성하여 클라우드로 전송한다. 미세조정 과정 개요도는 Fig. 3과 같다.

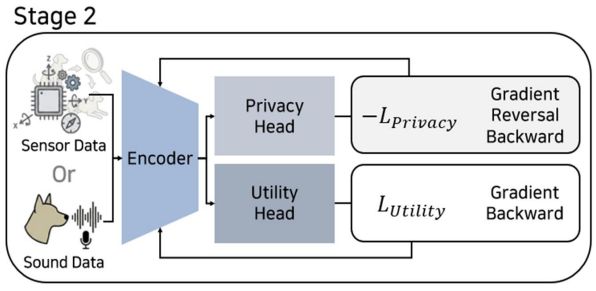

Fig. 3. Overview of fine-tuning process

3. Cloud-based inference

본 절에서는 엣지 디바이스로부터 전송된 익명화된 데이터를 기반으로 최종적인 상황을 판단하고 사용자에게 서비스를 제공하는 클라우드 서버의 역할을 기술한다. 클라우드 서버는 엣지 단계에서 생성된 잠재 특징 벡터를 수신하여 행동 인식 추론을 수행한다.

기존의 클라우드 중심 모델이 원본 영상이나 음성 데이터를 직접 분석하기 위해 고비용의 연산을 수행해야 했던 것과 달리, 제안 시스템의 서버는 이미 엣지 디바이스 단계에서 핵심 특징 추출과 정제가 완료된 잠재 특징 벡터를 입력받는다. 따라서 서버 측의 분석 모델은 복잡한 특징 추출기 없이, 단순한 완전 연결 계층으로 구성된 경량화된 분류기만으로도 높은 정확도의 추론이 가능하다.

보안 측면에서, 클라우드 서버에 저장되는 데이터는 원본 신호로의 역변환이 수학적으로 불가능한 잠재 특징 벡터 형태이다. 적대적 학습을 통해 개체 식별 정보가 소거되었으므로, 만약 클라우드 서버가 해킹되거나 데이터베이스가 유출되더라도 공격자는 해당 데이터가 어떤 사용자의 반려견인지 식별할 수 없다. 최종적으로 분석된 반려동물의 행동 상태(서다(Stand), 걷다(Walk), 앉다(Sit) 등)는 시계열 로그로 기록되며, 이상 행동 감지 시 사용자 애플리케이션으로 알림을 전송하는 모니터링 서비스로 연결된다.

IV. Experiment

1. Experimental environments and datasets

본 연구에서는 반려동물의 행동 분류 성능을 검증하기 위해 실제 환경에서 수집된 다중 모달리티 데이터셋을 구축하였다. 데이터 수집 장치는 IMU(ICM-20948)와 MCU (MAX32670GTL)를 사용하였으며, 이를 통해 100Hz의 샘플링 레이트로 센서 데이터를 확보하였다. 데이터 수집 시 디바이스는 모든 실험 개체에 대해 하네스를 활용하여 목 위치에 고정 부착되었다[20].

데이터의 신뢰성을 위해 모든 수집 과정은 카메라 녹화와 병행되었으며, 수집된 센서 데이터와 비디오 타임스탬프를 대조하여 라벨을 부여하는 검토 과정을 거쳤다. 수집된 원시 데이터 중 결측치가 발생한 구간에 대해서는 10% 미만의 결측에 한해 스플라인 보간(Spline Interpolation)을 적용하여 신호를 복원하였고, 10% 이상의 결측이 발생한 데이터는 분석에서 제외하였다. 정제된 데이터는 2초 단위의 윈도우로 분할되었으며, 윈도우 간의 겹침(Overlap)은 적용하지 않았다. 또한, 각 센서 축의 데이터 범위를 통일하고 학습 효율을 높이기 위해 모든 센서 데이터는 1에서 -1사이의 범위로 정규화(Normalization)를 수행하였다.

모델의 학습 및 일반화 성능 향상을 위해 세 가지 데이터 증강 기법을 적용하였다. 첫째, Time Reverse 기법을

통해 시계열 데이터의 순서를 역순으로 배열하여 시간 축을 반전시켰다. 둘째, Flip 기법을 사용하여 데이터 값을 x축 기준으로 반전시켜 패턴을 다양화하였다. 셋째, Jittering을 통해 데이터에 가우시안 노이즈를 추가함으로써 실제 수집 환경에서 발생할 수 있는 잡음에 대응하도록 하였다. 전처리가 진행된 데이터 시각화는 Fig 4와 같다.

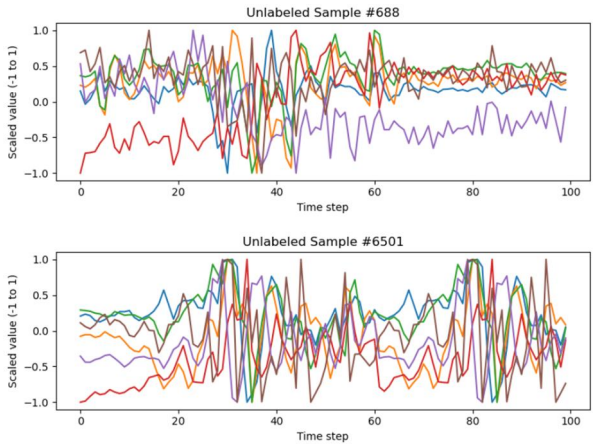

Fig. 4. Visualizing pet sensor datasets

실험에는 총 7마리의 반려견이 참여하였으며, 행동 클래스는 서다(Stand), 걷다(Walk), 앉다(Sit), 눕다(Lie), 먹다(Eat)의 5종으로 정의하였다. 데이터셋 분할은 세션 독립 방식을 채택하여 학습, 검증, 테스트 세트를 구성하였다. 반려동물 행동 데이터의 상세 규모는 Table 1과 같다.

Table 1. Dataset of pet behavior

Dataset	Label	Counts	Total
for pre-traing	-	16,152	16,152
for fine-tuning	Stand	1,584	6,979
	Walk	1,110	
	Sit	1,817	
	Lie	13,61	
	Eat	1,107	
for test	Stand	396	1,745
	Walk	278	
	Sit	454	
	Lie	340	
	Eat	277	

또한, 반려동물의 음성 데이터를 탐지 및 분류하기 위해, 공개된 벤치마크 데이터셋인 ArlingtonCL2/DogSpeak_Dataset과 일상 소음 데이터셋인 AI-Hub의 도시 소리 데이터를 융합하여 활용하였다. 먼저 사전 학습 단계에서는 도시 소리 데이터를 활용하여 모델이 일상생활에서 발생할 수 있는 다양한 음향적 특징과 배경 잡음에 대한 기반 지식을 학습하도록 한다. 이후 미세 조정 단계에서는 사전학습에 사용되지 않은 도시 소리 데이터와 반려동물 음성 데이터를 혼합하여, 복잡한 일상 소음 환경 속에서도 반려동물의 발성을 정밀하게 탐지할 수 있도록 모델을 최적화한다. 구축된 반려동물 음성 데이터의 상세 규모는 Table 2와 같다.

Table 2. Dataset of audio detection

Dataset	Counts
City noise for pre-training	61,778
City noise for fine-tuning	542,793
pet audio	77,197

반려동물 음성 데이터셋은 샘플레이트 16000으로 통일하였다. 또한 데이터셋에 포함된 개체 수는 총 156마리이며, 모든 오디오 데이터는 2초 길이로 분할하여 스펙트로그램으로 변환한다. 원본 오디오와 변환된 스펙트로그램의 시각화는 Fig. 5와 같다.

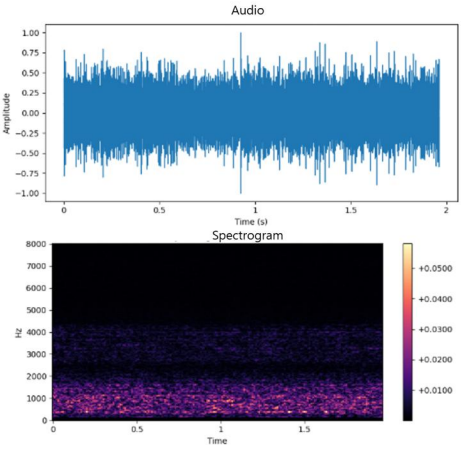

Fig. 5. Visualizing pet audio datasets

또한, 본 논문에서 제안하는 프라이버시 보존형 모니터링 시스템의 모든 모델 구현 및 실험은 PyTorch 딥러닝 프레임워크를 기반으로 수행되었다. 실험의 공정성과 재현 가능성을 확보하기 위해 사용된 구체적인 하드웨어 사양, 그리고 학습률(Learning Rate) 및 배치 크기(Batch Size)를 포함한 주요 하이퍼파라미터 설정은 Table 3과 같다.

Table 3. Experimental environments

Device	Version
GPU	RTX 3090
RAM	64GB
CUDA	12.6
Python	3.10
Pytorch	12.6
Batch Size	32

2. Experiment

2.1 Performance evaluation

본 절에서는 제안하는 프라이버시 보존형 MAE 모델이 실제 환경에서 수집된 데이터를 바탕으로 얼마나 정확하게 반려동물의 행동을 분류하고 음성을 탐지할 수 있는지, 즉 모델의 유용성을 객관적으로 검증하기 위해 수행한 비교 실험 결과를 기술한다. 실험의 신뢰성을 확보하기 위해, 시계열 데이터 및 오디오 신호 처리 분야에서 널리 활용되는 대표적인 딥러닝 모델들을 비교군으로 선정하여 동일한 조건 하에서 성능을 평가하였다.

본 연구의 목적은 제안하는 익명화 전략이 유틸리티-프라이버시 균형에 미치는 효과를 검증하는 데 있다. 이에 따라 비교군은 첫째, 시계열 모델의 대표 구조인 LSTM(Long Short-Term Memory), 둘째, 국소 패턴 추출에 강한 CNN, 셋째, 제안 모델과 동일한 계열이지만 자기지도 사전 학습을 적용하지 않은 트랜스포머(Transformer)이다. 이러한 설계는 모델 계열별 성능 차이뿐 아니라, MAE 기반 사전학습 및 적대적 프라이버시 학습이 실제로 성능 향상과 프라이버시 보호에 기여하는지를 공정하게 분리하여 검증하기 위한 것이다. 최신 대규모 모델들은 사전학습 데이터 규모, 파라미터 수, 연산 예산 차이로 인해 제안 기법 자체의 효과를 비교하기 어렵다는 점에서 본 연구의 직접 비교 대상으로는 포함하지 않았다. 또한 이를 통해 단순한 모델 구조의 차이뿐만 아니라, 본 연구의 핵심인 MAE 기반 자기지도 학습의 효과를 다각도로 분석하고자 하였다.

성능 평가는 모델의 전체적인 예측력을 나타내는 정확도(Accuracy)와, 각 클래스 간의 데이터 불균형 문제를 고려하여 정밀도(Precision)와 재현율(Recall)의 조화 평균을 계산한 F1-점수(Macro F1-score)를 주요 정량 지표로 채택하였다. 모든 모델은 동일한 전처리 과정과 하이퍼파라미터 설정을 적용하여 공정한 비교가 이루어지도록 설계하였으며, 실험 결과는 다음 Table 4와 같다.

Table 4. Performance Evaluation table

Task	Model	Acc	F1
Behavior Classification	CNN	0.889	0.890
	LSTM	0.844	0.831
	TF	0.878	0.867
	Base	0.903	0.894
Audio Detection	CNN	0.979	0.978
	LSTM	0.985	0.986
	TF	0.987	0.987
	Base	0.989	0.990

실험 결과를 구체적으로 분석해 보면, 제안하는 MAE 기반 모델이 IMU 센서 기반의 행동 분류와 반려동물 음성 탐지에서 모든 모달리티에서 기존 베이스라인 모델들을 상회하는 성능을 기록하였다. 먼저, 전통적인 시계열 처리 모델인 LSTM과 CNN은 데이터의 지역적인 패턴은 포착하였으나, 복잡한 행동 패턴이나 배경 소음이 혼재된 데이터의 전역적인 문맥을 파악하는 데에는 구조적인 한계를 보였다. 반면, 트랜스포머 기반 모델들은 어텐션 메커니즘을 통해 데이터의 긴 의존성을 효과적으로 학습함으로써 더 우수한 분류 성능을 입증하였다.

특히, 본 실험에서는 동일한 백본을 사용하는 표준 트랜스포머와 제안 모델 간의 성능 차이가 확인되었다. 제안 모델은 행동 분류에서 비교 모델 대비 약 2~3%, 음성 탐지에서 0.2~1% 높은 정확도를 달성하였다. 이는 지도 학습 방식이 정답 맞히기에만 급급해 과적합되기 쉬운 반면, 제안 모델은 MAE 기반의 마스킹 및 복원 사전 학습을 통해 데이터의 고유한 분포와 내재적 구조를 스스로 학습하여 더욱 견고한 특징 표현력을 갖추었기 때문이다.

또한, 현실적인 데이터 수집 환경을 고려할 때, 반려동물의 불규칙하고 다양한 행동 패턴에 일일이 정답 레이블을 부착하는 작업은 막대한 비용과 시간이 소요되는 과정이다. 제안하는 MAE 기반의 프레임워크는 레이블이 없는 대규모의 원시 데이터를 사전 학습에 활용함으로써 이러한 한계를 극복할 수 있다.

2.2 Anonymization evaluation

본 절에서는 제안하는 시스템이 행동 인식 성능을 유지하면서도, 민감한 개체 식별 정보를 얼마나 효과적으로 은닉할 수 있는지를 정량적 및 정성적으로 평가한다. 이를 검증하기 위해, 엣지 디바이스에서 클라우드 서버로 전송되는 잠재 특징 벡터를 가로채어 해당 데이터의 주인을 식별하려는 재식별 공격 시나리오를 가정하였다.

실험을 위해 별도의 공격자 모델을 설계하였다. 이 모델은 3층의 완전 연결 계층으로 구성된 분류기로, 추출된 잠재 특징 벡터를 입력받아 해당 데이터가 어떤 반려견으로부터 수집되었는지를 예측하도록 학습되었다. 평가는 방어 기법이 적용되지 않은 일반적인 MAE 베이스라인 모델과, 본 논문에서 제안하는 적대적 학습 기반의 익명화 모델을 비교하는 방식으로 진행되었다. 정량적 실험 결과표는 Table 5와 같다.

Table 5. Anonymization evaluation table

Task	Model	Utility		Privacy	
		Acc	F1	Acc	F1
Behavior Classification	Base	0.903	0.894	0.907	0.862
	Our	0.905	0.895	0.146	0.087
Barking Detection	Base	0.989	0.990	0.564	0.187
	Our	0.974	0.981	0.355	0.052

실험 결과, 방어 기법이 적용되지 않은 베이스라인 모델은 행동 인식 정확도 0.903만큼이나 높은 0.907의 개체 식별 정확도를 보였다. 또한 음성 탐지 정확도 0.989, 개체 식별 정확도 0.564를 보였다. 이는 딥러닝 모델이 학습 과정에서 행동 정보뿐만 아니라, 개체 고유의 걸음걸이, 체형, 음성 등 개체 식별 정보까지도 잠재 공간에 과도하게 보존하고 있음을 시사한다. 즉, 일반적인 특징 추출 방식은 데이터 유출 시 심각한 프라이버시 침해로 이어질 수 있음을 확인하였다.

반면, 제안하는 적대적 학습 기반 모델은 행동 인식 성능을 0.904로 준수하게 유지하면서도, 개체 식별 정확도를 0.149 수준까지 낮추었고, 음성 탐지 성능은 0.974로 준수하게 유지하면서도, 개체 식별 정확도를 0.355 수준까지 낮추었다. 이는 모델이 식별 정보를 전혀 학습하지 못하고 무작위로 정답을 추측하는 확률에 근접한 수치이다. 이러한 결과는 제안 시스템의 프라이버시 손실이 인코더로 하여금 개체 식별에 기여하는 특징만을 선택적으로 소거하고, 행동 분석에 필요한 정보만을 남기도록 성공적으로 유도했음을 입증한다.

다중 행동 분류를 진행한 데이터셋에 대해서 세부적인 클래스별 행동 분류 성능과 프라이버시 보존 효과를 시각적으로 분석하기 위해, 테스트 데이터셋에 대한 혼동 행렬(Confusion Matrix)을 도출하였으며 이는 Fig. 6과 같다. Fig. 4의 상단 행은 행동 인식, 하단 행은 개체 식별에 대한 결과를 나타내며, 좌측열은 베이스라인 모델, 우측열은 제안 모델의 결과이다.

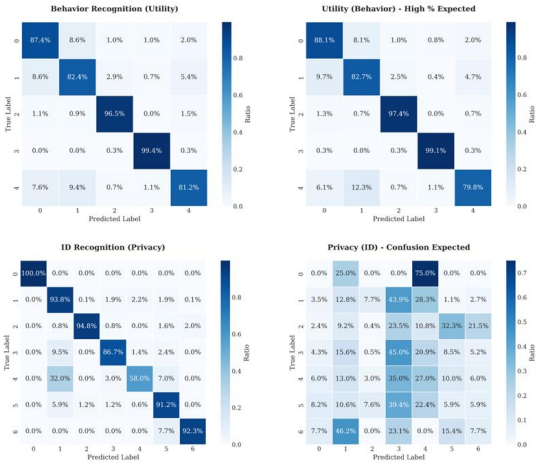

Fig. 6. Confusion matrix of behavior classification

우선 상단의 행동 인식 결과를 살펴보면, 제안 모델(우측 상단)은 프라이버시 보호 기법이 적용되었음에도 베이스라인 모델과 마찬가지로 대부분의 샘플이 정답을 의미하는 대각 성분에 집중되어 있음을 확인할 수 있다. 이는 행동 인식을 위한 결정 경계가 명확히 유지되고 있음을 의미한다.

반면, 베이스라인 모델(좌측 하단)은 개체 식별에서도 뚜렷한 대각선 분포를 보이며 훈련 데이터에 포함된 반려동물의 ID를 거의 완벽하게 구분해낸다. 그러나 제안 모델(우측 하단)의 경우, 대각선의 형태가 완전히 붕괴되었으며 예측값이 특정 클래스에 편중되거나 전반적으로 산개되는 양상을 보인다. 이는 인코더가 추출한 잠재 특징 벡터 내에서 개체를 구분할 수 있는 정보가 성공적으로 소거되었으며, 모델이 더 이상 특정 반려견을 식별하지 못하고 통계적 혼동 상태에 빠졌음을 시각적으로 확인하였다.

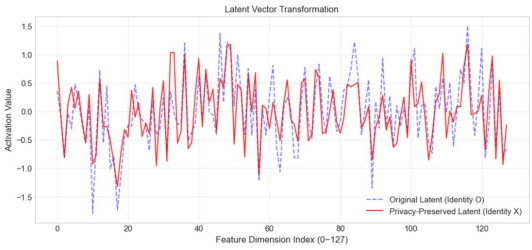

Fig. 7. Visualize latent vector comparisons

또한, 이러한 특징 소거 효과를 내부 구조 관점에서 분석하기 위해 Fig. 7과 같이 잠재 벡터의 차원별 활성화 값 분포를 시각화하였다. Fig. 7은 동일한 입력 데이터에 대해 베이스라인 모델(Original Latent, 파란 점선)과 제안 모델(Privacy-Preserved Latent, 빨간 실선)이 추출한 128차원 잠재 특징 벡터의 변화를 비교한 것이다.

그래프를 상세히 분석해보면, 두 벡터는 전체적인 파형의 등락과 같은 전역적인 흐름은 유사한 경향성을 보인다. 이는 제안 모델이 행동 인식에 필수적인 문맥 정보에 대한 정보는 훼손하지 않고 보존하고 있음을 시사한다.

그러나 국소적인 관점에서는 뚜렷한 차이가 관찰된다. 파란색 점선이 특정 차원(예: Index 10, 70, 95 부근)에서 강한 피크나 급격한 변동을 보이며 개체 고유의 특징을 드러내는 반면, 빨간색 실선은 해당 구간에서 값이 크게 억제되거나 반대 방향으로 이동하는 등 의도적인 편차를 보인다. 이는 적대적 학습 과정에서 프라이버시 손실이 인코더로 하여금 개체를 특정할 수 있는 미세 패턴만을 선택적으로 교란시켰음을 의미한다.

V. Conclusion

본 논문에서는 스마트 홈 환경에서 급증하는 반려동물 모니터링 서비스의 프라이버시 침해 위협을 원천적으로 차단하기 위해, 적대적 학습 기반의 프라이버시 보존형 엣지-클라우드 시스템을 제안하였다.

제안하는 시스템은 MAE를 백본으로 채택하여, 레이블이 부족한 현실적인 데이터 환경에서도 반려동물의 행동과 음성 신호에 내재된 문맥적 특징을 효과적으로 학습할 수 있었다. 특히, 엣지 디바이스 단계에서 수행되는 특징 추출 과정에 프라이버시 손실을 도입하여, 행동 분석에 필요한 유용성 정보는 최대화하고 개인을 특정할 수 있는 개체 식별 정보는 최소화하는 Min-Max 최적화를 수행하였다.

실험 결과, 제안 모델은 기존의 LSTM, CNN 및 트랜스포머 모델 대비 가장 우수한 행동 분류 및 음성 탐지 성능을 달성하였으며, 동시에 재식별 공격 시나리오에서 개체 식별 F1 점수를 무작위 추측 수준인 0.087(행동 분류), 0.052(음성 탐지)까지 낮추는 탁월한 보안 성능을 입증하였다. 또한, 잠재 특징 벡터의 시각화 분석을 통해 제안 모델이 행동 패턴의 전역적인 흐름은 보존하되 개체 고유의 신원 서명만을 선택적으로 소거함을 확인하였다. 이는 단순히 데이터를 암호화하는 기존 방식을 넘어, 딥러닝 모델의 잠재 공간 자체를 프라이버시 친화적으로 변환했다는 점에서 중요한 의의를 갖는다.

본 연구는 재식별 분류 성능 저하를 중심으로 수행되었다. 또한 제안 프레임워크의 실용성은 구조적으로 엣지-클라우드 분할에 적합함을 보였으나, 실제 엣지 디바이스에서의 지연 시간, 메모리 사용량, 전력 소모, 잠재벡터 전송량을 정량적으로 측정하지 못한 한계가 있다. 향수 연구에서는 다양한 위협 모델에 대한 다면적 프라이버시 평가와 실제 디바이스 기반 시스템 레벨 벤치마크를 수행할 예정이다. 이를 위해 모델의 연산 복잡도를 더욱 최적화하여 초저전력 MCU(Micro Controller Unit) 수준의 극소형 엣지 디바이스에서도 실시간 추론이 가능하도록 모델 경량화(Quantization & Pruning) 연구를 진행할 계획이다. 아울러, 현재의 IMU 센서 및 음성 데이터를 넘어 비디오 데이터를 포함한 멀티모달 환경에서의 프라이버시 보존 기법으로 본 프레임워크를 확장 및 검증할 것이다.

REFERENCES

[1] P. Pico-Valencia and J. A. Holgado-Terriza, "The Internet of Things Empowering the Internet of Pets—An Outlook from the Academic and Scientific Experience," Applied Sciences, vol. 15, No. 4, 2025.

[2] P. Kasnesis, V. Doulgerakis, D. Uzunidis, D. G. Kogias, S. I. Funcia, M. B. González, C. Giannousis and C. Z. Patrikakis, "Deep learning empowered wearable-based behavior recognition for search and rescue dogs," Sensors Vol. 22, No. 3, 2022.

[3] N. B. Nguyen, K. Chandrasegaran, M. Abdollahzadeh and N. M. Cheung, "Re-thinking model inversion attacks against deep neural networks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16384-16393, 2023.

[4] J. R. Gómez-Armenta, H. Pérez-Espinosa, J. A. Fernández-Zepeda and V. Reyes-Meza, Automatic classification of dog barking using deep learning. Behavioural Processes, 2024

[5] A. Abzaliev, H. Perez-Espinosa and R. Mihalcea "Towards dog bark decoding: Leveraging human speech processing for automated bark classification," Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). , pp. 16480-16486, 2024.

[6] J. Jeong, M. Cho, P. Benz and T Kim, "Noisy adversarial representation learning for effective and efficient image obfuscation," Uncertainty in Artificial Intelligence, PMLR, pp. 953-962, 2023.

[7] A. Dosovitskiy and T. Brox, "Inverting visual representations with convolutional networks," CVPR, pp. 4829-4837, 2016.

[8] N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramer, B. Balle, D. Ippolito and E. Wallace, "Extracting training data from diffusion models," 32nd USENIX security symposium (USENIX Security 23), 2023.

[9] J. Q. Zhang, M. Xiong, P. Li, J. Yuan and H. Zhu, "Attribute Inference Attacks in Edge-Cloud Collaborative Inference," in IEEE Transactions on Information Forensics and Security, 2024.

[10] D. Pasquini, G. Ateniese and M. Bernaschi, "Unleashing the

Tiger: Inference Attacks on Split Learning," CCS, pp. 2113-2129, 2021.

[11] E. Erdoğan, A. Küpçü and A. E. Çiçek, "Unsplit: Data-oblivious model inversion, model stealing, and label inference attacks against split learning," Proceedings of the 21st Workshop on Privacy in the Electronic Society, 2022.

[12] M. Gong, K. Pan, Y. Xie, A. K. Qin and Z. Tang "Preserving differential privacy in deep neural networks with relevance-based adaptive noise imposition." Neural Networks 125, pp. 131-141, 2020.

[13] P. Vepakomma, A. Singh, O. Gupta and R. Raskar "NoPeek: Information leakage reduction to share activations in distributed deep learning," 2020 International Conference on Data Mining Workshops (ICDMW). IEEE, 2020.

[14] Y. Kang, J. Hauswald, C. Gao, A. Rovinski, T. Mudge, J Mars and L. Tang, "Neurosurgeon: Collaborative intelligence between the cloud and mobile edge," ACM SIGARCH, Vol. 45, No. 1, pp. 615-629 2017.

[15] D. Yao, "Towards Privacy-Preserving Split Learning for ControlNet," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2025.

[16] V. Prodomo, R. Gonzalez and M. Gramaglia "SCIPER: Secure Collaborative Inference via Privacy-Enhancing Regularization," in IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2024.

[17] H. Wang, C. Qiu, C. Zhang, J. Xu and C. Su, "P-CA: Privacy-Preserving Convolutional Autoencoder-Based Edge-Cloud Framework," in Mathematics, Vol. 12, No. 16, 2024.

[18] P. Kumpulainen, A. V. Cardó, S. Somppi, H. Törnqvist, H. Väätäjä, P. Majaranta, Yulia Gizatdinova, C. H. Antink, V. Surakka, M. V. Kujala, O. Vainio and A. Vehkaoja, "Dog behaviour classification with movement sensors placed on the harness and the collar." Applied Animal Behaviour Science 241, 2021

[19] R. D. Chambers, N. C. Yoder, A. B. Carson, C. Junge, D. E. Allen, L. M. Prescott, S. Bradley, G. Wymore, K. Lloyd and S. Lyle, "Deep learning classification of canine behavior using a single collar-mounted accelerometer: Real-world validation," Animals, Vol. 11, No. 6, 2021

[20] Kim JinAh, Hyungju Kim, Chan Park and Nammee Moon. “Deep Learning-based Pet Monitoring System and Activity Recognition device.” Journal of The Korea Society of Computer and Information, 27(2), 25-32, 2022

Authors

Hyuksoon Choi received the B.S. degree in the School of Computer Engineering from Hoseo University, Korea, in 2024, and the M.S. degree in the Department of Computer Science from Hoseo University, Korea, in 2026.

Hyuksoon Choi is currently with the Department of Computer Science at Hoseo University, Korea. His research interests include self-supervised learning, contrastive learning, time-series data, and big data processing and analysis.

JinHwan Yang received the B.S. degree in the School of Computer Engineering from Hoseo University, Korea, in 2024, and the M.S. degree in the Department of Computer Science from Hoseo University, Korea, in 2026.

JinHwan Yang is currently with the Department of Computer Science at Hoseo University, Korea. His research interests include unsupervised learning, multimodal embedding, restoration learning, and audio processing.

Nammee Moon received the B.S., M.S., and Ph.D. degrees in Computer Science and Engineering from Ewha Womans University, Korea, in 1985, 1987, and 1998, respectively. Dr. Moon joined the faculty of the Department of Computer Science and Engineering at Hoseo University, Korea, in 2008. She is currently a Professor in the Department of Computer Science and Engineering, Hoseo University. Her research interests include social learning, human-computer interaction (HCI), user-centric data, and big data processing and analysis.