

UNIK 프레임워크를 활용한 고양이 행동 인식 연구[†]

(Cat Pose Recognition with UNIK Framework)

최 연 서 [‡]
(Yeonseo Choi)

이 수 진 [§]
(Sujin Lee)

요 약 1인 가구의 수가 증가함과 동시에 반려묘를 포함한 다양한 반려동물을 키우는 가구가 늘어나고 있다. 이에 따라 반려동물을 대상으로 다양한 기술적 시도들이 일어나고 있다. 이러한 기술적 시도에는 딥러닝 기반으로 다양한 동물의 행동을 인식하는 기술 또한 포함되어 있다. 하지만 특정 동물에게만 적용하는 기술은 반려견 대상이 대부분이다. 그에 반해 반려묘 즉, 고양이에게만 특화된 딥러닝 기반 모델들은 아직 많은 시도가 이루어지지 않았다. 본 논문에서는 이러한 문제점을 해결하기 위해 키포인트 정보만을 이용하여 고양이에게 특화된 딥러닝 모델을 제안하고자 한다. 기존에 구현되어있는 스켈레톤 기반 행동 인식 딥러닝 모델 중 하나인 UNIK를 활용하여 고양이의 행동을 인식하도록 모델을 학습한다.

키워드 : 행동 인식, 고양이, UNIK, 딥러닝

Abstract As the number of single-person households increases, the number of households with various companion animal, including cats, is increasing. As a result, various technological attempts have been made on companion animal. These technological attempts include deep learning based pose recognitions of various animals. However, most of the technologies that apply only to certain animals are for dogs. By contrast, deep learning-based models specializing in cats, have yet to be tried. In order to solve this problem, this paper proposes a deep learning model specializing in cats using key points. We adopt the UNIK framework, the most promising skeleton-based action recognition framework based deep learning techniques to train our cat pose recognition model.

Key words : action recognize, pose, cat, UNIK, deep learning

1. 서 론

반려인 1500만 시대가 도래했다. 기존 저출산·고령화

등으로 1인 가구 증가 및 소득이 증대함에 따라 반려가구가 늘어난 것으로 보인다. 2021년 6월 4일에 KB금융지주 경영연구소가 발표한 ‘2021 한국반려동물보고서’[1]에 따르면 한국에서 반려동물을 기르는 반려가구는 640만 가구로 전체 가구의 24.7%를 차지하는 것으로 드러났다.

반려가구의 증가에 비례하여 반려동물을 대상으로 하는 다양한 기술적인 시도 또한 늘어났다. 이는 현재 주목 받는 딥러닝 또한 마찬가지로, 동물 동작 인식 관련 연구가 많이 진행되고 있다. 그러나 반려견이 대부분이며 반려묘를 대상으로 하는 연구는 상대적으로 매우 제한적이다.

본 연구에서는 기존에 구현된 스켈레톤 기반 행동 인식 모델인 UNIK[2]를 활용하여 고양이의 행동을 인식함으로써, 제한된 키포인트로 고양이의 행동을 11가지로 분류하는 모델을 제안한다.

[†] 본 연구는 2004년도 정보통신부 지원 대학 IT 연구센터 육성지원사업(C1090-0403-0005)의 연구비 지원으로 수행하였습니다.

[‡] 학생회원 : 성균관대학교 정보통신공학부
hjikim@dosan.skku.ac.kr

[§] 비 회 원 : 성균관대학교 정보통신공학부
yeosh72@hanmail.net

[¶] 종신회원 : 성균관대학교 정보통신공학부 교수
dhwon@dosan.skku.ac.kr

논문접수 : 2003년 10월 27일

심사완료 : 2004년 09월 02일

Copyright©2004 한국정보과학회 : 개인 목적이나 교육 목적의 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 정보통신 제31권 제6호(2004.12)

2. 관련 연구

2.1 고양이 행동 인식 모델

본 논문의 주제인 고양이 행동 인식과 관련된 최신의 선행 연구를 살펴보면 다음과 같다. Yoona Choi 등[3]은 반려묘의 모니터링과 질병 진단 시스템을 위하여 시공간적 특성을 학습하는데 효과적인 능력을 가지고 있는 LSTM기반 모델과 탈부착식 센서인 비컨을 활용한 시스템을 제안하였고, Jun-Hee Lee 등[4]은 객체 검출을 위한 모델 중 하나인 YOLO와 영상의 RGB 프레임과 Optical Flow 프레임 정보를 동시에 사용하는 Two-Stream YOLO를 이용하여 실시간으로 고양이 행동인식을 하는 시스템을 제안하여 걷거나 눕는 등의 행동에 대한 7가지 클래스에 대해서 고양이 행동을 분류하였다. 하지만 이러한 7가지의 행동보다 세부적인 행동에 대해서 분류하는 모델은 연구가 크게 이루어지지 않았다.

2.2 스켈레톤 기반 행동 인식

스켈레톤 기반 행동 인식하는 방법들은 영상 속에 나타난 사람의 자세들을 프레임별로 나타내고 이러한 자세들의 변화를 통해 행동을 인식하는 방법을 의미한다.

Hyun-Jae Bae 등[5]은 키포인트를 추출하고 사람의 행동 인식 목적으로 구축한 LSTM 기반의 모델을 제안하였다. Maosen Li 등[6]은 A-line 추론 모듈이라는 인코더-디코더 구조를 도입하여 동작별 잠재 의존성을 파악하는 AS-GCN 기반의 모델을 제안하였다.

전술한 바와 같이 스켈레톤 기반 및 키포인트를 이용하여 행동 인식을 한 모델은 GCN 방식과 LSTM 방식을 사용한다. 본 논문에서는 GCN 방식을 보완한 UNIK 모델을 이용해 고양이에 대해 행동 인식을 진행하였다.

3. UNIK : 스켈레톤 기반 동작 인식

기존의 방식인 GCN(Graph Convolutional Network)은 사전에 정의된 사람의 토폴로지에 의존하는 골격에 대한 특징을 효과적으로 추출할 수 있는 기법이다. 그러나 이와 같은 방법은 다른 사람의 위상 구조로 일반화하기 힘들다는 단점이 존재한다. 본 논문에서 사용하는 스켈레톤 기반 행동 인식 모델인 UNIK 같은 경우 기존의 GCN에서 사용되는 인접행렬 대신 키포인트간의 의존성을 더 잘 나타낼수 있는 의존행렬(Dependency Matrix)로 대체하여 학습을 진행한다. 기존의 인접행렬 같은 경우 인접한 키포인트와의 영향을 주로 다루어 멀리 있는 키포인트에 대해서 영향을 덜 받았지만, 의존행렬 같은 경우 인접했다는 정보가 아닌 두 키포인트 간의 의존도가 얼마나 큰가에 대해서 학습을 하기 때문에 멀리 있는 키포인트에 대해서도 영향을 받는다. 고양이가 같은 경우 유연한 몸을 가지고 있으므로 멀리 있는 키포인트 간의 영향이 사람에 비해서 크다고 판단하였기에 UNIK를 사용하였다.

4. Data Augmentation

4.1 반려동물 구분을 위한 동물 영상 데이터셋

표 1. 12가지의 행동 클래스의 각 갯수
Table 1. The number of each of the 12 action classes.

Pose	Training set	Validation set
grooming(그루밍하는 동작)	5184	648
tailing(꼬리 흔드는 동작)	3106	389
armstretch(팔을 뻗어 휘적거리는 동작)	2948	369
walkrun(걷거나 달리는 동작)	2576	323
laydown(옆으로 눕는 동작)	1734	217
sitdown(발을 숨기고 웅크리고 앉는 동작)	1470	184
getdown(납작 엎드리는 동작)	1105	139
lying(배를 보여주는 동작)	940	118
heading(머리를 돌리는 동작)	865	109
footpush(앞발로 꼭꼭 누르는 동작)	770	97
roll(좌우로 뒹구는 동작)	666	84
arch(허리를 아치로 세우는 동작)	180	23
합계	21544	2700

한국정보화진흥원(NIA) 주관으로 구축한 AI 기술 및 제품 서비스 인프라인 AI-Hub의 데이터[7]를 활용하였다. 반려동물 구분을 위한 동물 영상 데이터셋의 구축 년도는 2020년이며, 2021년 10월 7일자로 추가 개방된 데이터셋까지 사용했다. 최소 10초~최대 60초 길이의 영상 파일로 수집되었으며 연구에서 사용한 고양이 데이터는 총 학습 데이터가 21544개, 총 테스트 데이터는 2700개다.

한국정보화진흥원(NIA) 주관으로 고양이를 대상으로 12가지의 행동과 함께 관절의 15개 키포인트 값 X, Y으로 이루어진 반려동물 영상 데이터셋이 구축되어 있다.

4.2 클래스 및 키포인트 개수 변경

기존의 데이터셋을 학습시켜본 결과 laydown(옆으로 눕는 동작)과 lying(배를 보여주는 동작)를 구분해줄 수 있는 키포인트가 없다고 판단해 고양이의 배 부위를 나타내기 위한 키포인트를 추가하여 총 16가지의 키포인트를 사용하였다.

행동은 표 1과 같이 총 12가지의 클래스가 기본적으로 구분되어있으나, 영상을 분석한 결과 납작 엎드린 동작(getdown)과 발을 숨기고 웅크리고 앉는 동작(sitdown)의 차이가 꼬리가 보이는지 여부만 차이가 났기 때문에 클래스를 별도로 처리할 필요가 없다고 판단해 이 두 개의 클래스를 합쳐 총 11가지의 클래스를 16개의 좌표에 따라 인식하도록 학습을 진행했다.

4.3 Data Augmentation

표 2. 시도 전후 11가지 행동 클래스 별 Training set 갯수

Table 2. The number of Training sets by 11 action classes before and after the Data Augmentation..

Pose	Before	After
grooming(그루밍하는 동작)	5184	4874
tailng(꼬리 흔드는 동작)	3106	2968
armstretch(팔을 뻗어 휘적거리는 동작)	2948	2829
walkrun(걸거나 달리는 동작)	2576	2464
laydown(옆으로 눕는 동작)	1734	2254
sitdown(발을 숨기고 웅크리고 앉는 동작)	1470	3699
getdown(납작 엎드리는 동작)	1105	2254
lying(배를 보여주는 동작)	940	2254
heading(머리를 들이대는 동작)	865	2254
footpush(앞발로 꼭꼭 누르는 동작)	770	2254
roll(좌우로 뒹구는 동작)	666	2254
arch(허리를 아치로 세우는 동작)	180	2254
합계	21544	30358

4.2에서 언급한 대로 학습을 진행한 결과 데이터 불균형문제가 있다고 판단을 하였고 이 문제를 해결하기 위해서 데이터 증강을 수행하였다. 좌우 반전, 회전, SMOTE를 사용했다. SMOTE 기법은 데이터셋의 Null 키포인트를 채우는 것으로 진행했다. 데이터 증강을 수행한 후 데이터의 개수는 표2와 같다.

5. 성능 평가

5.1 세부 파라미터

적합한 계산을 위하여 세부 파라미터를 수정했다. 데이터셋에 맞추어 Joint를 16개로 설정하고 Num Class를 11개로 설정했다. In Channels은 2, Num Heads는 {3}, Optimizer는 SGD로 설정했다. Bacth Size는 8, Epoch은 30, Num Worker는 4로 설정했다.

5.2 실험 결과 및 평가

표 3. Data Augmentation 적용 후 정확도

Table 3. Accuracy after applying Data Augmentation.

실험 내용	Top1	Top5
50도 회전, 좌우 반전	72.20%	97.45%
SMOTE	70.59%	97.52%
SMOTE, 좌우 반전	73.25%	97.28%
SMOTE, 좌우 반전, 90도 회전	76.09%	98.11%

데이터셋에 주어진 대로 15개의 키포인트로 12가지 행동에 대하여 학습을 진행했을 때의 Top1이 73.89%, Top5는 97.48%이다. 여기에 키포인트를 더해 총 16개의 키포인트로 11가지 행동에 대하여 학습을 진행했을 때 Top1은 75.37%, Top5는 97.56%이다. Top1 기준 정확도

가 1.48% 향상되었다.

키포인트와 행동 클래스를 변경한 데이터셋에 Data augmentation을 적용한 결과는 표3과 같다. SMOTE, 좌우 반전, 90도 회전을 적용했을 때 Top1은 76.09%, Top5는 98.11%이다. 키포인트와 행동 클래스만 변경했을 때와 비교했을 때 Top1 기준 0.72% 향상되었다.

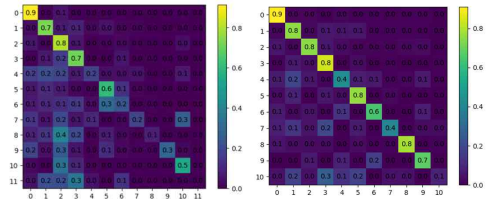


그림 1. 학습 1

그림 2. 학습 2

Fig. 1. Training 1

Fig. 2. Training 2

그림 1은 데이터셋에 주어진 대로 15개의 키포인트로 12가지의 행동에 대하여 학습을 진행했을 때 각 클래스의 예측률을 나타낸 것이다. 그림 2는 키포인트 및 행동 클래스 변경하고 Data Augmentation을 적용하고 학습을 진행했을 때 각 클래스의 예측률을 나타낸 것이다. 처음 성능이 낮았던 행동 클래스의 정확도가 상승한 것을 확인할 수 있다.

정확도가 가장 높았던 실험 결과를 기준으로 했을 때 성능이 0.5 미만으로 낮은 클래스는 4번 클래스인 laydown(옆으로 눕는 동작), 7번 클래스인 laying(배를 보여주는 동작), 10번 클래스인 arch(허리를 아치로 세우는 동작)로 각각 0.4, 0.4, 0.1이었다.

각 클래스가 정확도가 낮은 이유를 분석해본 결과는 다음과 같다. 배를 보여주는 동작 같은 경우 해당 클래스로 분류하는 조건은 배를 보이는 동작이 1회이상 포함이거나 배를 보인 상태에서 몸의 움직임이 있을 것이다. 이때 해당 조건상 전자 같은 경우 좌우로 뒹구는 동작(roll)과 유사한 면이 있고 후자 같은 경우 다른 동작과 키포인트 상에서 크게 특출난 점이 없기 때문이다. 옆으로 눕는 동작 또한 키포인트가 다른 행동과 비교를 했을 때 큰 차이가 없다. 허리를 아치로 세우는 동작같은 경우 기본 데이터의 양이 적기 때문에 이를 데이터 증강을 이용해 데이터의 수를 늘렸음에도 불구하고 한계가 있을 것이며 해당 행동의 경우 중요하다고 생각되는 허리의 위치를 알 수 없고 고양이의 배 부분에 해당하는 키포인트도 다리의 시작점 4개에 대해서 평균을 이용한 구했기 때문에 해당 행동에 대해서는 정확한 배 부분에 해당하는 위치가 아닐 가능성이 크기 때문으로 분석하고 있다.

6. 결론 및 향후 과제

본 연구에서는 고양이의 행동 인식을 목적으로 키포인

트 데이터를 스켈레톤 기반 행동 인식 모델인 UNIK로 학습하여 11개의 고양이 행동 클래스에 대하여 sota 성능을 확보하였다. 이때 가장 좋은 성능을 내기 위해 SMOTE와 회전 90도 및 좌우 반전을 적용하였다.

향후과제로 성능을 잘 내지 못한 클래스에 대해서도 성능을 올리기 위해 키포인트를 추가하거나 테스트를 위한 데이터를 제작하기 위해 Pose Estimation을 이용한 키포인트 추출모델을 만드는 것이다. 또한 UNIK가 아닌 다른 프레임워크에 대해서도 고양이 행동 분류를 학습시키고 이에 대해서 성능을 비교하는 것이다.

참 고 문 헌

- [1] Influences such as the increase in the number of ‘Pes-kog-jog’ and single-person households: “Light and Shadow” in the era of 15 million partners [in the issue],
<https://m.segye.com/view/20210603519756> [accessed: Oct. 29, 2021]
- [2] Di Yang, Yaohui Wang, Antitza Dantcheva, Lorenzo Garattoni, Gianpiero Francesca, and Francois Bremond, “UNIK : A Unified Framework for Real-world Skeleton-based Action Recognition”, Jul. 2021.
- [3] Yoona Choi, Heechan Chae, Jonguk Lee, Daihee Park, and Yongwha Chung, “Cat Monitoring and Disease Diagnosis System based on Deep Learning”, Vol. 24, No. 2, pp. 233-244, Feb. 2021.
- [4] Jun-Hee Lee, Jonguk Lee, Yoona Choi, Daihee Park, and Yongwha Chung, “Real-Time Cat Behavior Recognition System using Two-Stream YOLO”, pp. 408-411, May. 2019.
- [5] Hyun-Jae Bae, Gyn-Jin Jang, Young-Hun Kim, Jin-Pyung Kim, “LSTM(Long Short-Term Memory)-Based Abnormal Behavior Recognition Using AlphaPose”, Vol.10, No. 5, pp.187-194, Mar. 2021.
- [6] Maosen Li, Siheng Chean, Xu Chen, Ya Zhang, Yanfeng Wang, and Qi Tian, “Actional-Structural Graph Convolutional Networks for Skeleton-based Action Recognition”, Apr. 2019.
- [7] animal image for companion animal classification, <https://aihub.or.kr/aidata/34146>, [accessed: Oct. 11, 2021]
- [8] ANIMAL-POSE DATASET, <https://sites.google.com/view/animal-pose/>, [accessed: Oct. 30, 2021]