

Two-Stream YOLO 를 이용한 실시간 고양이 행동 인식

이준희* 이종욱**[†], 최윤아*, 박대희**, 정용화**

*고려대학교 컴퓨터정보학과

**고려대학교 융합소프트웨어학과

e-mail : watrtdc@korea.ac.kr

Real-Time Cat Behavior Recognition System using Two-Stream YOLO

Jun-Hee Lee*, Jonguk Lee**[†], Yoona Choi*, Daihee Park**, Yongwha Chung**

*Dept. of Computer Information Science, Korea University

**Dept. of Computer Convergence Software, Korea University

요 약

고양이를 기르는 가구의 증가와 함께 건강한 애묘 방법을 찾는 애묘인 또한 증가하고 있다. 본 논문에서는 고양이의 건강 상태를 모니터링하기 위해 반드시 선행되어야만 하는 고양이의 행동 정보를 딥러닝 방법론을 기반으로 인식하고자 한다. 인식을 위해 먼저, 카메라 센서를 이용하여 고양이 영상 데이터를 수집한 후, 수집된 영상에서 RGB 프레임과 optical flow 프레임 정보를 각각 수집한다. 각각의 프레임은 RGB Network 와 Flow Network 에 입력되고, 두 네트워크 결과 정보에 대하여 concatenation 을 수행한다. 연계된 특징 정보는 행동 인식 알고리즘인 Two-Stream YOLO 에 입력되어 고양이의 행동을 인식한다. 고양이의 행동 인식은 일곱 개의 클래스로 나누어 진행하였다. 행동 인식 실험 수행 결과 mAP 와 fl-score 모두에서 0.9 이상의 높은 성능을 보였으며, 실시간으로 수행 가능함을 확인하였다.

1. 서론

시장 조사 서비스를 제공하는 APPA(American Pet Products Association)에 따르면, 전 세계 반려동물 시장에서 미국 시장이 압도적이며, 미국인의 약 68%가 반려동물을 기르는 것으로 조사되었다. 그 중 고양이의 개체 수가 물고기 다음으로 가장 많은 것으로 조사되었다. 또한 미국 반려동물 산업은 1994 년부터 꾸준히 성장하고 있으며, 2018 년에는 약 80 조에 이르는 지출이 발생할 것으로 예상하였다[1]. 국내의 경우 2012 년부터 2017 년까지 반려묘를 기르는 가구의 증가율은 같은 기간 반려견보다 약 6 배 이상 높다고 보고되었다[2]. 이처럼 반려 동물로서의 고양이에 대한 관심이 매우 증가하고 있으며, 이와 함께 건강하게 고양이를 키우는 방법에 관한 관심도 증가하고 있다. 그러나, 상대적으로 고양이의 건강 상태를 모니터링하기 위한 연구의 수는 매우 적다[3].

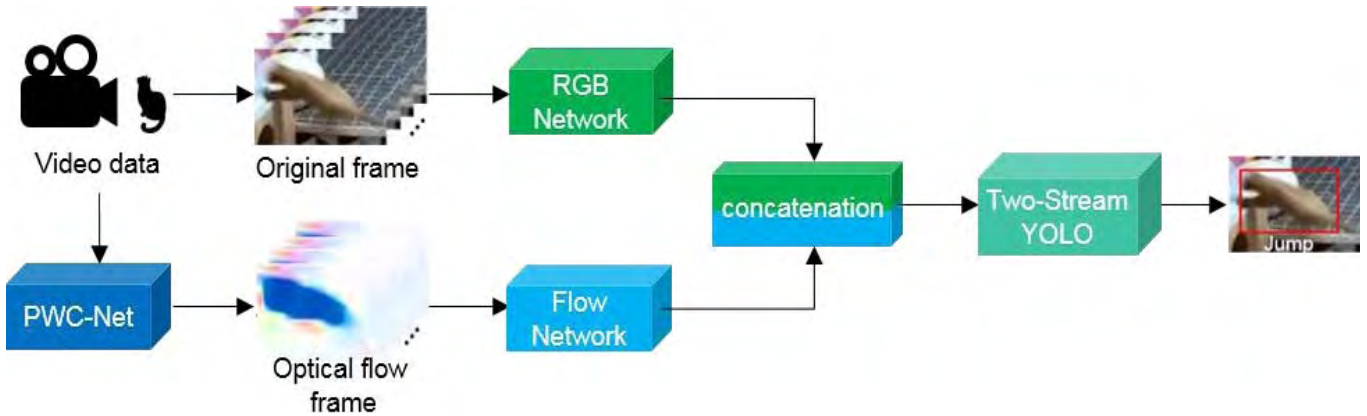
고양이의 건강과 관련한 질병들 중에서 전염병을 제외하고 가장 많은 부분을 차지하는 질환은 요로계 관련 질병이며, 높은 비율로 사망에 까지 이르기도 한다. 이와 같은 요로계 질병을 발생 초기에 발견하기 위해서는 고양이의 배뇨 행동을 24 시간 자세하게 관찰하고 기록해야 한다[4]. 그러나, 고양이의 행동을

애묘인이 하루 종일 세세하게 관찰한다는 것은 불가능한 부분이다. 본 연구에서는 이와 같은 고양이의 건강 상태를 모니터링하기 위해 반드시 필요한 행동 정보를 딥러닝 알고리즘을 기반으로 인식하는 방안을 제안한다.

최근 이미지 분류에서 매우 놀라운 성과를 보인 CNN(Convolutional Neural Network) 알고리즘을 동영상으로 확장하여, 해당 영상에 나오는 객체에 대한 행동 인식이 가능함을 확인하는 연구들이 발표되었으며[5], CNN 구조를 개선하여 행동을 효과적으로 인식하는 YOLO(You Look Only Once)와 같은 방법을 기반으로 행동 인식 성능을 개선하려는 연구들도 보고되었다[6-7]. 일반적으로 이미지 분류에서는 RGB 픽셀 값으로 되어있는 외관(appearance) 정보가 유일한 단서이지만, 비디오에서는 영상 내 객체가 시간의 흐름에 따라 변화하는 움직임(motion) 정보도 부가적으로 취득할 수 있다. 이를 활용하여 영상에 나타나는 외관 정보는 RGB 를 활용하고, 시각적 움직임 정보는 optical flow 를 이용하여 취득한 후 두 정보를 연계하는 Two-Stream Network 이 제안되었으며, RGB 만을 이용한 경우보다 객체의 행동 인식 성능이 향상 되었음이 보고 되었다[8].

본 논문에서는 Ali 등[8]의 구조를 참조하여 비디오

[†] 교신저자



(그림 1) 고양이 행동 인식을 위한 Two-Stream YOLO 구조

를 통해 들어오는 RGB 기반의 외관 정보와 optical flow 기반의 움직임 정보를 연계한 후, 연계된 정보를 YOLO 에 입력하는 Two-Stream YOLO 구조를 제안하며, 이를 기반으로 고양이 행동을 인식하는 시스템을 제안한다.

2. Two-Stream Network 기반 고양이 행동 인식 시스템

본 연구에서는 고양이를 촬영한 영상에서 고양이의 행동을 인식하기 위하여 다음과 같은 과정을 수행한다. 1) 취득한 영상에서 RGB 정보와 optical flow 정보를 각각 획득한다. 2) 객체의 외관 정보가 포함된 RGB 와 움직임 정보가 내재된 optical flow 를 연계(concatenation) 한다. 3) 연계된 정보를 Two-Stream YOLO 에 입력하여 고양이의 행동을 인식한다. (그림 1 참조)

2.1 PWC(Pyramid, Warping, and Cost Volume)-Net

본 연구에서는 고양이를 촬영한 영상에서 움직임 정보를 취득하기 위한 방법으로 optical flow 정보를 이용하고자 한다. 이 때 다양한 optical flow 취득 알고리즘들 중 최근에 제안된 PWC-Net[9]을 활용한다. PWC-Net 은 이전의 optical flow 추출 알고리즘과 비교하여 낮은 end-point error(EPE)를 보여주는 데, 이때 사용된 EPE 는 추정된 optical flow 벡터 V_{est} 와 정답지 optical flow 벡터 V_{gt} 의 차이 값에 의해 계산된다. 또한, PWC-Net 은 GPU(Graphic Processing Unit)를 이용한 병렬 연산이 가능하기 때문에 영상의 실시간 처리가 가능하며, 행동이 발생한 영역의 optical flow 정보를 효과적으로 추출하는 장점이 있다. 이와 같은 장점을 반영하여 PWC-Net 를 활용하여 optical flow 정보를 추출한다.

2.2 YOLOv2

객체의 행동 인식을 위해 제안된 YOLO 는 CNN 을 기반으로 만들어졌으며, 이미지 내의 경계상자(bounding box)와 클래스 확률(class probability)를 하나의 회귀 문제로 간주하기 때문에, 객체의 종류와 위

치를 동시에 탐지할 수 있다. YOLO 네트워크는 경계상자 탐지와 클래스 분류를 동시에 수행하는 end-to-end 구조를 가졌으며, 실시간으로 수행이 가능하다. 또한, YOLO 는 인식하고자 하는 객체 뿐만 아니라 전체 이미지를 학습하기 때문에 이미지에서 제공하는 문맥을 보다 잘 이해할 수 있다.

YOLOv2 에서는 이전 버전의 문제점인 작은 물체의 탐지 및 인식 성능 향상을 위하여, pass-through 레이어를 추가하여 초기 레이어에 있는 정보를 마지막 레이어에 추가하여 성능을 향상시켰다. 부가적으로, 출력 계층과의 연결을 fully-connected 에서 convolutional connect 로 대체하여, 분류에 입력되는 특징 벡터의 파라미터의 개수를 감소시켰다. 또한, YOLO 는 fully-connected 를 convolutional connect 로 대체로 인한 성능 저하를 줄이기 위하여 anchor box 라는 개념을 도입하였다. 이와 같은 노력에 의해 YOLOv2 는 기존 YOLO 보다 처리 속도를 개선시킴과 동시에 객체 인식 성능도 안정적으로 유지하는 장점이 있다[6-7].

2.3 Two-Stream YOLO

기존의 행동 인식 연구들은 두 가지 다른 정보를 갖는 스트림을 이용하는 경우, 해당 스트림을 각각 학습시킨 후 개별 스트림으로부터 나온 결과를 퓨전(fusion) 알고리즘을 기반으로 연계한 후 인식 알고리즘 결과를 도출하였다. 그러나, 두 개의 스트림을 따로 학습 시키는 것은 외관 신호와 움직임 신호 사이의 해당 시간 차원에서의 높은 연관 정보를 사용할 수 없게 된다. 본 연구는 RGB Network 와 Flow Network 의 마지막 레이어에서 concatenation 을 진행함에 따라 외관과 움직임 신호 사이에 시간 차원의 연관 정보를 유지한다. RGB Network 와 Flow Network 는 YOLOv2 을 기반으로 설계되었으며, concatenation 된 값은 Two-Stream YOLO 의 입력으로 사용된다. Two-Stream YOLO 는 입력된 값을 이용하여 이미지에 해당되는 경계상자, 클래스 그리고 선택된 클래스의 확률 값을 결과로 생성한다.

3. 실험 및 결과 분석

3.1 고양이 영상 데이터 취득 환경

실험을 위한 데이터는 라즈베리파이에 연결된 카메라를 이용하여 10fps의 속도로 저장하였다. 저장된 영상 데이터는 전체 7개의 클래스로 구분하였으며, 각각의 데이터는 standing, seating, walking, drinking, eating, jumping, lying으로 나누어 저장하였다. 각 클래스에 대한 예는 그림 2와 같다.



(그림 2) 클래스에 따른 프레임 예

개별 클래스는 500 프레임, 총 3500 프레임이며, 8:2의 비율로 나누어 각각 학습 및 테스트 데이터로 사용하였다. 실험에 사용된 컴퓨터의 사양은 Ubuntu 16.04 운영체제, Intel i7-6700k 프로세서와 32GB 메모리, 그리고 GTX1080 그래픽 카드이다. 학습 시 학습율 0.00005, momentum 0.9, decay 0.0005로 설정하였으며, 활성화 함수는 leaky ReLU를 사용하여 진행하였다.

3.2 Two-Stream YOLO을 이용한 행동 인식 결과

표 1은 본 연구에서 제안한 고양이 행동 인식 시스템의 성능을 평가한 실험 결과이다. 본 연구에서 제안하는 고양이 행동 인식 시스템의 성능을 평가를 위해 먼저, 객체의 영역을 탐지한 후 고양이의 행동 인식 수행하게 된다. 이 때, 실제 객체가 있는 영역과 제안한 시스템에 의해 객체 영역으로 판단된 영역의 겹쳐짐 정도(intersection over union)를 α 로 표현한다. 겹쳐진 영역이 20%이면 α 는 0.2, 50%이상 겹쳐져 있으면 0.5로 수치를 설정하여, 해당 객체의 탐지 성능을 평가하는 사전 기준이 된다. 또한, 고양이 행동 인식 성능을 정량적으로 평가하기 위하여, mAP(mean Average Precision)와 f1-score를 채택하였다. F1-score의 경우 recall과 precision의 조화 평균으로 계산되며,

mAP는 recall과 precision을 그래프로 표현한 precision-recall 그래프의 평균으로 계산된다. 두 수치 모두 1에 가까울수록 좋은 성능을 의미한다. α 값을 나누어 성능을 평가한 결과 f1-score와 mAP 모두에서 0.9 이상의 우수한 인식 성능을 보임을 확인할 수 있으며, 한 장의 이미지 처리 속도는 0.052 초로 비디오 취득 시간이 10fps임을 고려해볼 때, 실시간으로 고양이의 행동 인식이 가능하다는 사실을 실험적으로 검증하였다.

<표 1> 고양이 행동 인식 시스템 실험 결과

mAP ($\alpha = 0.2$)	mAP ($\alpha = 0.5$)	F1-score ($\alpha = 0.2$)	F1-score ($\alpha = 0.5$)
0.966	0.906	0.965	0.908

4. 결론

본 논문은 고양이의 건강 상태를 모니터링하기 위한 사전 단계인 행동 인식을 딥러닝 기반 알고리즘을 활용하여 수행한 연구이다. 제안한 Two-Stream YOLO 구조를 기반으로 고양이의 기본적인 행동을 실시간으로 식별하였으며, 실험 결과 높은 인식률을 확인하였다.

감사의 글

본 연구는 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018R1D1A3B07044938)

참고문헌

- [1] APPA, "Pet Industry Market Size & Ownership Statistics," [https://www.americanpetproducts.org/press_industrytrends.asp]
- [2] 강규민, "유로모니터 '올해 韓 펫케어 시장 1.6조원까지 성장'", 파이낸셜뉴스 [http://www.fnnews.com]
- [3] 이유정, 윤민희, 변진우, 이민희, 강원국, 권혜영, 이윤성, 김옥진, "고양이 문제 행동의 예방과 교정법에 대한 고찰", 한국동물매개심리치료학회지, Vol. 3, No. 1, pp. 79-83, 2014.
- [4] P. Johnson-Bennett, "Think Like a Cat: How to Raise a Well-Adjusted Cat--not a Sour Puss," Penguin, 2011
- [5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "Ssd: Single Shot Multibox Detector," European Conference on Computer Vision, pp. 21-37, 2016.
- [6] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263-7271, 2017.
- [7] 이준희, 채희찬, 이종욱, 박대희, 정용화, "YOLO를 이용한 개별 돼지 탐지 시스템," 한국정보처리학회 추계학술대회, pp. 724-726, 2017.
- [8] A. Ali and G. W. Taylor, "Real-Time End-to-End Action Detection with Two-Stream Networks," Conference on

Computer and Robot Vision (CRV) IEEE, Vol. 15, pp. 31-38, 2018.

- [9] D. Sun, X. Yang, M. Y. Liu, and J. Kautz, "Pwc-net: Cnns for Optical Flow using Pyramid, Warping, and Cost Volume," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8934-8943, 2018.