

# 최근 정보통신 기술 딥러닝과 Object Detection 과 Skeleton Pruning 을 통한 반려견 행동 파악 서비스에 관한 연구

고현경, 박현정, 서윤경, 이영기  
아주대학교

kohy0329@ajou.ac.kr., jhikm1003@ajou.ac.kr , carys3115@ajou.ac.kr , t9rres@ajou.ac.kr

## A Study on the Deep Learning, Object Detection and Skeleton Pruning for Detecting action of dogs systems

Koh Hyeon Gyeong, Park Hyeon Jung, Seo Yun Kyung, Lee Yeoung Ki  
Ajou Univ.

### 요 약

본 논문에서는 딥러닝, Object Detection 과 Skeleton Pruning 을 이용하여 반려견의 행동을 파악하는 기술에 대하여 말하고자 한다. 반려견의 행동을 모두 다 아는 사람들은 드물다. 그렇기 때문에 사람들은 검색엔진을 이용한다. 다만, 검색엔진을 이용할 때에 한계점인 언어의 모호성과 정보의 부정확성이 존재한다. 이러한 한계점을 사진촬영을 이용한 검색엔진과 비슷하게 만들고자 한다. 학습 시킬 사진을 가지고 YOLO 와 GrabCut 알고리즘을 통해 배경제거를 한 후, 배경제거 된 사진을 Skeleton Pruning 을 이용하여 뼈대와 그에 대한 좌표를 추출한 후, Deep Learning 을 통하여 뼈대에 대한 정보들을 학습 시킨 후, 후에 다시 사진이 입력될 시 학습된 데이터를 가지고 반려견의 행동을 판단 가능하게 해주는 기술을 연구하였다.

### I. 서론

근래, 반려견을 키우는 사람들의 수는 늘어나고 있다. 하지만 반려견을 키우는 모든 사람들은 반려견이 어떤 행동을 취하여 어떤 말을 전하고자 하는지 모두 알 수는 없다. 그렇기 때문에 사람들은 구글 또는 네이버와 같은 검색엔진을 이용하여 언어를 이용하여 검색을 하며, 또는 사진과 같이 올리며 다른 사람들이 온라인을 통하여 진단하여 주기 까지를 기다린다.

하지만 언어를 통한 검색은 언어의 모호성으로 인해 얻고자 하는 정보에 대하여 한계를 줄 수 있으며, 또한 검색을 하여도 정확한 정보를 얻을 수 있다는 확신은 없다. 사진과 같은 부가 정보와 같이 글을 게시하여도 이에 대한 답변을 기다리기 위해서는 시간이 필요하다.

이러한 문제점들을 해결하기 위해서 Object Detection, Grabcut, Skeleton Pruning 그리고 Deep Learning 을 이용하여 반려견의 행동을 알아보고자 한다.

사진 촬영을 하면 주어진 사진에서 Object Detection 을 이용하여 반려견을 인식한 뒤, 반려견을 제외한 그 외의 배경요소들을 제거한다. 그 후, 배경이 제거된 사진을 가지고 Skeleton Pruning 을 통하여 뼈대와 각 좌표들에 대한 정보들을 추출한다. 이를 통하여 Deep Learning 된 학습 Data 와 비교하여 현재 반려견의 상태가 어떠한지 판단하고자 한다.

### II. 본론

#### 1. GrabCut 및 Object Detection 기술을 사용한 배경제거

GrabCut 알고리즘은 사진에 전경 객체를 포함하는 직사각형을 지정해주면 전경 객체를 추출해주는 알고리즘이다. 직사각형을 정해 주면 직사각형 중앙에 추출하고자 하는 전경 객체가 존재한다고 가정하여 전경 객체의 색상 분포와 배경의 색상 분포를 추정하여 이를 통해 전경을 추출해낸다. 전경, 배경 부분을 사용자가

보조적으로 직접 입력해주면 더 좋은 결과를 얻을 수 있다. 그런데 어플리케이션 사용자에게 사진 촬영 이후 이러한 입력들을 해달라고 요구할 수 없다. 따라서 직사각형 지정, 전경, 배경 부분 입력은 자동화 하여야 한다. 직사각형 지정은 Object Detection 기술을 사용하여 해결한다. Object Detection 기술을 사용한 YOLO 는 사진 내에 물체를 감지하면 이 물체가 무엇인지, 어디에 있는지 알려준다. 이를 사용해 사진 내에 반려견이 존재하는지 파악할 수 있을 뿐만 아니라 물체가 어디에 있는지 직사각형으로 알려주므로 이 직사각형을 GrabCut 에 사용할 수 있다. 다음으로 전경, 배경 부분을 보조적으로 입력하는 것을 해결하기 위해 사진에 Median filter 를 적용한다. Median filter 를 적용하면 전경 객체들은 전경 객체의 주된 색상으로 뭉게지고 배경도 배경의 주된 색상으로 뭉게진다. 이렇게 하면 전경 객체, 배경의 색상 분포가 뚜렷해지므로 전경, 배경 부분을 직접 입력해주는 것과 비슷한 효과를 얻을 수 있다.

#### 2. Skeleton Pruning

Skeleton Pruning 은 그림의 Outline 을 통하여 그림의 뼈대를 구하는 기술이다. Skeleton Pruning Algorithm 은 총 4 가지 단계로 이루어진다. 1. Contour approximation, 2. Skeleton Computation, 3. Pruning, 그리고 4. Skeleton Vectorization 이다.

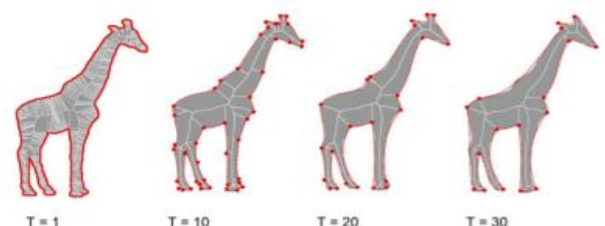
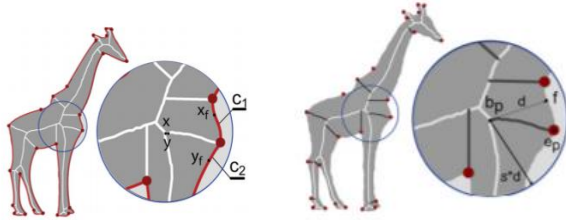


Figure 1

Grabcut 을 통하여 주어진 사진으로 Bezier curve 를 이용하여 contouring 을 진행한다. 이 때, 정해진 vertices 들의 수 대로 테두리를 잡고 curve 와 pixel 의 거리차가 threshold 보다 크다면 curve 를 split 해 나가면서 contouring 을 한다. Threshold 로 정해진 vertices 들의 수에 대하여 잔가지가 뺄어 나와 메인 뼈대를 이룬다.

불완전한 메인 뼈대를 이루면, contouring line 과 메인 뼈대 사이에 존재하는 잔 뼈대들을 제거해준다.



**Figure 2**

두 번째 단계에서는 Figure 2의 왼쪽과 같이 뼈대 점 X는 동일한 거리 d에 있는 다른 곡선에 속하는 가장 가까운 경계 점을 가진 이웃 점 Y가 있는 경우 선택됩니다. 그렇다면 이는 메인 뼈대 포인트로 지정되는 것이다.

Final Pruning 을 해주는 단계에서는 Figure 2의 오른쪽과 같이 두 번째 메인 뼈대와 잔 뼈대를 구분해주는 것을 토대로  $|e_p - b_p|_2 \leq s * |f - b_p|_2$  식을 이용하여 잔 뼈대와 필요없는 contouring 으로 point 들도 삭제해준다.

그 후, 마지막으로 두 번째 단계에서 선택된 X와 같은 좌표들을 기준으로 vector 화 시켜준다. 그렇다면 다음과 같은 결과 값을 도출해 낸다.

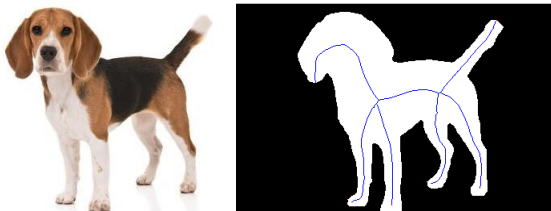
### 3. Deep Learning

딥러닝 기술은 많은 부분에서 사용된다. 먼저 추출된 뼈대를 정확하게 활용하려면 이 뼈대의 머리가 어디인지 파악하여야 하는데, 여기에 딥러닝 기술을 적용한다. face detection에 최적화 되어있는 dlib 라이브러리를 사용하여 반려견 머리를 학습시켜 딥러닝 모델을 만든다. 머리를 찾아내면 자세 학습에 사용하기 위하여 뼈대를 머리를 기준으로하여 정렬한다.

반려견의 자세가 무엇을 의미하는지 알기 위해서도 딥러닝 기술을 적용한다. 정렬된 뼈대 데이터들을 사용하는 딥러닝 모델을 제작한다. 손실함수로 입력이 0을 넘으면 그 입력을 그대로 출력하고 그 이하면 0을 출력하는 ReLU 함수를 사용한다. 출력값은 softmax 함수를 사용하여 가장 일치하는 자세를 찾아내도록 한다. 학습 방법은 정답이 1이고 나머지가 0인 One-hot encoding 방식과 무작위로 데이터의 일부만 학습하는 mini-batch 학습 방식을 사용한다. 경사 하강법을 이용하여 자세를 추론하기 위한 최적의 매개변수를 찾는다.

### III. 결론

본 논문에서는 위의 알고리즘들을 활용하여 실제 사진을 넣으면 뼈대 사진을 도출해내는 방법을 고안해 냈다.



**Figure 3**

위의 사진처럼 실제 강아지 사진을 입력 할 시, Object Detection으로 반려견이 어디에 존재하는지 detect한 후, 해당 object를 제외한 background를 제거한다. 그 후, 제거된 사진에서 skeleton pruning의 4단계를 이용하여, 메인 skeleton, end, 그리고 branch point를 추출할 수 있다.



**Figure 4**

하지만 Figure 5에서 볼 수 있듯이 추출한 뼈대만을 가지고 학습시키기에는 뼈대에서 어느 쪽이 꼬리인지 머리인지 헷갈릴 수 있기에 dlib을 이용하여 머리와 꼬리를 구분하도록 한다. dlib에서 머리를 파악하게 되는 bbox를 알려준다면 bbox의 (left-right의 middle point, top-bottom의 middle point)를 bbox의 중심좌표로 잡으며 branch, end point 중 가장 가까운 점을 머리 점으로 잡고 뼈대의 기준을 잡아주어 뼈대를 학습시키는데에 도움이 되도록한다.

dlib을 통하여 알아낸 bbox와 skeleton의 좌표들과 매칭을 시켜준다면 각 branch, end point에 해당 점들이 어느 부위를 나타내는지 라벨링을 시켜준다. 각 라벨링된 점들로 이루어진 뼈대들을 어느 자세의 뼈대인지 딥러닝을 통하여 학습을 시켜준다.

대부분의 뼈대 추출 알고리즘은 사람에 관해서만 존재하였다면, 위와 같은 여러 오픈소스들을 활용하여 반려견의 뼈대 추출을 가능하게 하였다. 또한 반려견이 아닌 반려묘 등의 동물들에서도 뼈대 추출이 가능할 것이다. 그런 경우, 본 논문에서 말하고 있는 기술은 동물과 관련된 일을 하는 경우라면, 다양한 측면에서 사용이 가능하다.

### ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW 중심대학지원사업의 연구결과로 수행되었음"(2015-0-00908)

### 참고 문헌

- [1] Andre's Soli's Montero n, Jochen Lang, 2001, Skeleton pruning by contour approximation and the integer medial axis transform, School of Electrical Engineering and Computer Science (EECS)
- [2] Joseph Redmon, Ali Farhadi, 2017, YOLO9000 Better, Faster, Stronger, Best Paper Honorable Mention
- [3] Carsten Rother Vladimir Kolmogorov Andrew Blak, "GrabCut" — Interactive Foreground Extraction using Iterated Graph Cuts, Microsoft Research Cambridge, UK
- [4] Robert Laganière, 2011, OpenCV 2 Computer Vision Application Programming Cookbook