

<1>

안녕하세요, 저희는 이번 여름세션에 고양이 감정분석으로 컴퓨터비전 프로젝트를 진행한 권민지, 김영홍, 서준형, 유민서, 김윤영입니다. 저는 발표를 맡은 14기 서준형입니다.

<2>

목차는 다음과 같습니다.

<3> 제목 슬라이드: 01 프로젝트 주제 선정 및 목표 설정

<4>

혹시 반려동물을 키우시는 분들이 있나요? 키우는 분들 중 대부분은 반려동물이 과연 무슨 생각을 하는지 궁금해하셨을것 같습니다. 이런 사소한 궁금증에서부터 프로젝트 주제를 선정하게 되었습니다. 기존에 존재하는 유머 앱에서는 감정을 랜덤으로 출력하는 방식으로 이루어진 것들이 많았는데, 이를 딥러닝의 computer vision 기술과 결합하여 문제를 풀어가하고자 하였습니다. 특히 사람 감정 분석 데이터셋은 많았는데 반려동물에서 그 수가 적다는 점에서도 프로젝트의 의의가 충분히 존재한다는 생각에 주제를 선정하였던 것도 있었습니다.

<5> 제목 슬라이드: 02 데이터셋

저희는 이번 프로젝트를 크게 두 방법으로 진행을 하였습니다. 이를 위한 데이터셋을 먼저 소개해드리겠습니다

<6>

yolo를 사용한 모델의 경우 roboflow에서 바운딩 박스가 라벨링된 이미지를 약 900장을 이용했고 분류 모델 사용시에는 이미지와 감정으로 이루어진 다른 데이터셋을 케글에서 탐색하여 활용했습니다.

<7>

이어서 keypoint에서 사용한 데이터셋은 두가지입니다.

첫 번째는 "CatFLW"라는 데이터셋으로, 다양한 환경과 조건에서 촬영된 고양이 얼굴 약 2000개의 이미지이며 고양이 얼굴의 48개의 주요 랜드마크와 얼굴에 대한 바운딩 박스를 라벨링 되어진 것이 특징입니다.

두번째는 "Cat Emotions"라는 데이터셋으로, 고양이의 다양한 감정을 분류하기 위한 약 700개의 이미지들이며 고양이의 감정을 나타내는 라벨이 부여되어 있습니다.

<8> 제목 슬라이드: 03 모델 소개

저희 프로젝트에서 사용한 모델들에 관하여 설명드리겠습니다

<9>

감정 분류에 사용된 모델은 다음과 같습니다. VGG, ResNet, DenseNet, Efficient Net 을 사용해 보았습니다 모두 분류 모델로 유명한 모델이므로 설명은 피피티를 참고해주 시면 좋겠습니다

<10>

Yolo란 You Only Look Once의 준말로 이미지를 한 번 보는 것으로 Object Detection을 수행하는 컴퓨터비전 알고리즘을 의미합니다. 흔히 코로나 시기에 체온측 정을 위해 카메라에 얼굴을 비추면 얼굴 영역을 감지하고 체온을 측정해준 기기가 대표 적인 활용 예시인데요, 이러한 알고리즘을 저희는 동물의 얼굴 영역을 detection하는데 에 이용했습니다. 프로젝트에서는 v8을 사용하였지만, 기초적인 이해를 위해 v1을 기반 으로 설명하겠습니다.

올로 모델은 이미지를 입력받은 뒤, $S \times S$ 크기로 나누어 grid cell을 생성합니다. 이후 각각의 cell별로 class 확률을 예측하고 bbox와 confidence score를 예측하는 동시에 cell별로 클래스 확률값을 예측합니다. 수식적인 부분은 시간상 설명을 생략합니다.

<11>

다른 cnn모델과 같이 올로도 최종 feature map을 생성하며 이를 위한 구조로 DarkNet을 설계합니다

<12>

이어서 얼굴 표정 감지, 자세 탐지, 활동인식 등 다양한 분야에 사용되는 keypoint detection의 가장 기본이 되는 모델인 keypoint rcnn에 대해 설명드리겠습니다. Keypoint R-CNN은 Mask R-CNN을 변형하여 사람의 신체 부위의 키포인트를 감지하 는 기능을 추가한 것입니다.

구체적으로 먼저, 이미지에서 객체를 감지하기 위해 Region Proposal Network(RPN) 를 사용하여 관심 객체가 있을 가능성이 높은 위치를 예측합니다. 이후, ROI-Align 기 법을 통해 이 영역을 고정된 크기의 feature map으로 변환하는 Mask R-CNN 과정을 통해 객체의 클래스와 위치를 예측합니다.

그리고 Keypoint R-CNN은 이 구조에 Keypoint Head라는 브랜치를 추가하여, 각 사 람의 신체 부위별로 키포인트를 예측합니다.

<13> 제목 슬라이드: 04 모델 구축 과정 1

저희 프로젝트의 두가지 방법론 중 첫 번째 방법인 YOLO와 Efficient Net 모델 구축

과정에 대하여 설명 드리겠습니다

<14>

코드는 크게 3가지 단계로 이루어집니다. 첫번째로, 얼굴 영역을 추출하기 위해 yolo v8 m 모델로 바운딩 박스를 잡는 작업을 진행했습니다. 기존 데이터셋에서 감정 라벨링을 지우기 위해 yaml(얌)파일을 수정한 뒤 데이터셋을 50 epoch동안 학습했습니다. 학습 세팅값은 모두 디폴트값으로 설정하였습니다.

<15>

두번째로 고양이의 이미지와 감정 라벨링이 되어있는 데이터셋을 대상으로 cv2 모듈을 이용해서 예측한 바운딩 박스의 좌표값대로 영역을 추출하여 잘라내고, 분류기에 사용할 수 있도록 저장하는 작업을 진행했습니다.

<16>

셋번째로, EfficientNet으로 감정을 행복, 우울, 혐오, 놀람, 화남, 두려움, 기타의 7가지로 분류하는 모델을 훈련했습니다. 이미지 개수가 부족했기에 다양한 증강을 학습데이터에 적용하였고, 첫 시도에서 valid 성과가 좋지 못하여 수기로 구글링한 이미지를 valid 데이터셋에 추가하여 학습했습니다. train 약 500개, valid 약 200개로, 크로스엔트로피 손실과 아담 옵티마이저를 이용하고 학습률 스케줄링을 통해 과적합을 최대한 방지하였습니다. 성능 비교를 위해 ResNet과 Vision Transformer를 사용하였으나 EfficientNet에서 가장 좋은 성능을 기록하였습니다. 측정한 성능은 화면의 그래프와 같습니다.

<17>

최종적으로 이 모델로 확인한 결과는 다음과 같습니다. 이미지를 입력하면 얼굴 영역만 자르고 예측한 감정을 확인할 수 있습니다.

<18> 제목 슬라이드: 04 모델 구축 과정 2

저희 프로젝트의 두가지 방법론 중 두 번째 방법인 Keypoint 모델 구축 과정에 대하여 설명 드리겠습니다

<19>

먼저 파이프라인입니다. 데이터셋은 동물 이미지에 키포인트가 라벨링된 catflw 데이터, 그리고 감정 라벨이 포함된 Cat Emotions 데이터입니다.

첫 번째 단계에서는 동물 얼굴의 키포인트를 추출할 모델을 학습합니다. 동물 얼굴을

추출할 수 있는 키포인트 모델이 따로 없기 때문에, 사람 얼굴 키포인트를 추출하는 모델을 동물 이미지와 감정 라벨 데이터를 사용해 새롭게 학습시킵니다.

두 번째 단계에서는 이 학습된 모델을 활용해 감정 라벨이 붙은 동물 이미지에 키포인트 라벨을 생성하여 기존 데이터셋에 추가합니다. 이렇게 생성된 키포인트 라벨은 JSON 파일로 저장하고, 이미지와 함께 시각화하여 잘 생성되었는지 확인합니다.

<20>

이어서 세 번째 단계에서는 이미지와 생성된 키포인트를 입력으로 받아서 감정을 분류하는 멀티모달 분류 모델을 학습하고

마지막으로 네 번째 단계에서는 새로운 동물 이미지가 들어왔을 때, 첫번째 단계에서 학습한 모델로 키포인트를 추출하고, 이 키포인트와 이미지를 세번째 단계에서 학습한 감정 분류 모델에 넣어서 최종적으로 감정을 분석하게 됩니다.

<21>

첫번째 단계에서 동물 얼굴의 키포인트를 추출할 모델을 재학습 시키기 위해 총 다섯가지의 pretrain 실험을 진행했습니다. 순서대로 설명하자면,

첫 번째 모델은 Rescale, `Normalize, ToTensor와 같은 기본 전처리 기법만 사용했고

두 번째 모델은 RandomCrop과 RandomHorizontalFlip 기법을 추가했습니다.

세 번째 모델은 OpenCV의 `CascadeClassifier를 활용해 얼굴을 검출했고

네 번째 모델은 dlib 라이브러리를 사용해 학습했고

마지막으로 다섯 번째 모델은 CascadeClassifier와 여러 데이터 증강 기법을 사용한 모델입니다.

<22>

해당 pretrain의 결과는 다음과 같습니다.

<23>

이어서 감정 라벨이 붙은 동물 이미지에 키포인트 라벨을 생성하는 인퍼런스 결과를 살펴보겠습니다. 왼쪽 이미지를 보면 키포인트가 위로 치우쳐 생성되었고, 오른쪽 이미지에서는 키포인트가 왼쪽으로 치우쳐 있는 것을 확인할 수 있습니다.

이렇게 비교적 아쉬운 결과가 나왔는데, 이를 보완하기 위해서는 학습 데이터의 양과 다양성을 늘려 모델이 다양한 얼굴 각도와 표정을 잘 반영할 수 있도록 하고 데이터 전처리 과정과 모델의 하이퍼파라미터 조정을 통해 키포인트의 정확도를 높이는 작업을 진행한다면 더 좋은 성능을 보일 것이라고 생각합니다.

<24> 제목 슬라이드: 05 프로젝트 한계 및 의의

<25>

크게 yolo와 keypoint 2가지의 모델을 통해 이목구비를 잡고 분류기를 연결한 2가지 방법에서의 모두 정확도가 다소 떨어졌고 이는 데이터셋의 부족에서 기인했습니다.]이에 많은 양의 이미지가 주어지고, 또한 더 다양하게 모델을 튜닝한다면 yolo모델, keypoint 모델 모두 한계를 극복할 수 있다고 느꼈습니다.

추후에 저희 모델이 충분한 데이터로 성능이 보장 된다면 다양한 동물에 적용 또한 가능할 것이라고 제안합니다.

<26>

지금까지 cv조 발표였습니다. 감사합니다.