

KT AIVLE School

3차 미니프로젝트_팀 발표 템플릿

AI 모델링 Report 올레TV 코너 몰아보기

AI 08반 25조

데이터 분석 및 학습 데이터 생성

동영상을 보았을때 코너마다 고정적으로 나오는것이 무엇일까?

✓ 상단코너



상단코너 이미지예시

✓ 고정출연진



고정출연진 이미지예시

전처리 과정

✓ 전처리 1 : train 데이터 생성

→ 학습하는데 필요없다고 생각되는 오프닝과 엔딩을 제외, 코너별로 파일 새로 생성

✓ 전처리 2 : 상단 코너 이미지 데이터 생성

→ 상단에 코너 이름이 고정적으로 위치하여 코너 이름이미지를 사용하여 모델링 하면 정확도가 높게 나올거라고 기대 함.

✓ 전처리 3 : 고정 출연진의 얼굴만 추출 - landmark 사용

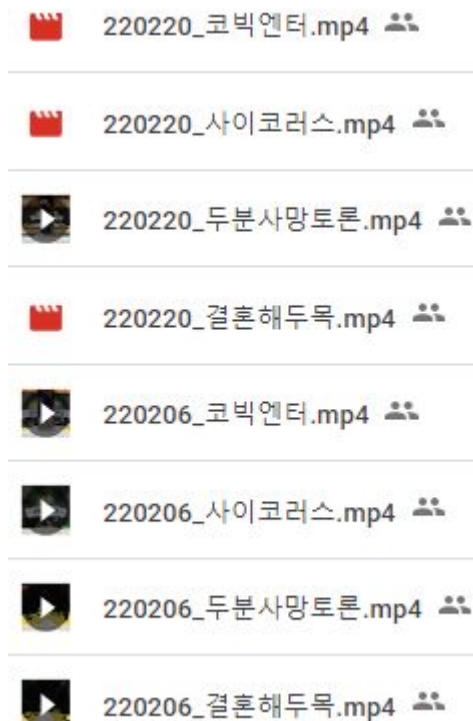
→ 코너마다 고정적으로 나오는 출연진이 있으므로 해당 이미지를 추출해 낼 수 있다면 좋은 결과를 가져올 수 있을 것이라고 기대 함.

✓ 전처리 4 : 상단 코너 이미지 데이터 + 고정출연진의 얼굴 데이터

→ 고정 출연진이 있을때 얼굴로 검출, 없을때 로고로 검출하여 전처리2,3 방식을 같이 사용한다면 좋은 결과를 가져올 수 있을 것이라고 기대함.

전처리 1 : train 비디오 데이터 전처리

- 오프닝, 엔딩 편집(삭제)
- 코너별 비디오 파일 생성



전처리 2 : 상단 코너 이미지 데이터 생성

- 비디오 프레임 5초 단위로 이미지 저장
- 코너별 비디오의 이미지 데이터 생성
 - 사이코러스 144개
 - 코빅엔터 177개
 - 결혼해두목 149개
 - 두분사망토론 156개
- 이미지 특정 위치의 **bounding box**로 **cropping**(잘라내기)



```
bounding_box = {  
    'start_x' : 40,  
    'start_y' : 60,  
    'end_x' : 157,  
    'end_y' : 75  
}
```



전처리 3 : 고정 출연진 얼굴만 추출

- 각 영상 별 1초 단위로 이미지 데이터를 추출.
- 복장과 머리스타일은 매코너 변경될 수도 있다고 판단.
=> 사람 고유의 특징인 얼굴을 이용하여 학습데이터로 구성.
- 50분의 동영상을 1초단위의 프레임으로 끊다보니 대량의 데이터가 생산됨.
- **landmark(face detection)**를 이용하여 얼굴 특징점을 추출.

얼굴 랜드마크(facial landmark)란?

- 얼굴의 랜드마크는 코, 눈, 입, 눈썹, 턱과 같은 것을 얼굴의 랜드마크라고 부른다.
- 얼굴 이미지 위에서 랜드마크라고 생각되는 곳들에 총 68개의 좌표를 찍는 방식으로 작동



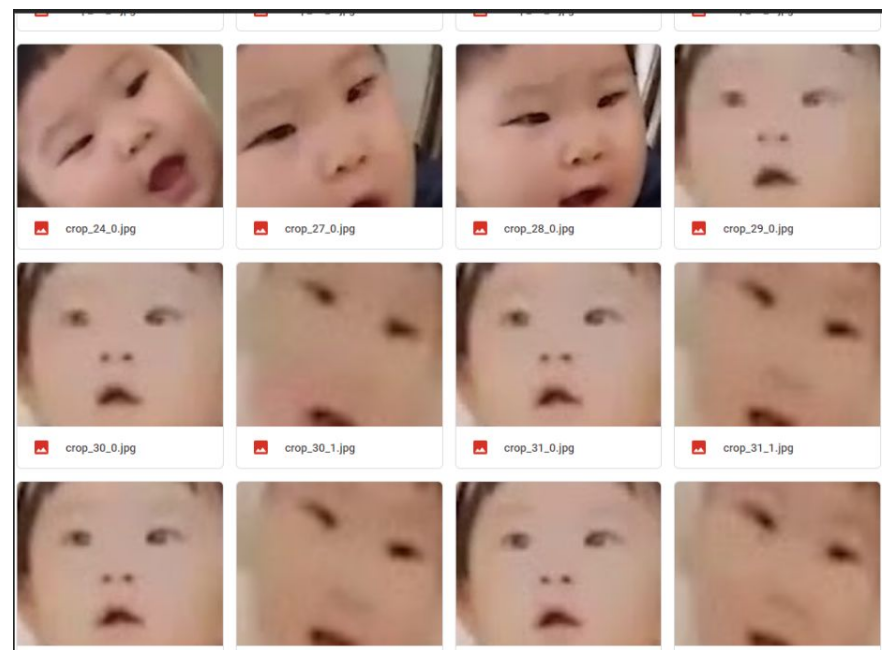
Facial landmark 실행 예시

얼굴 랜드마크(facial landmark)란?

■아래와 같이 얻은 x, y, w, h 좌표를 이용해서 image crop



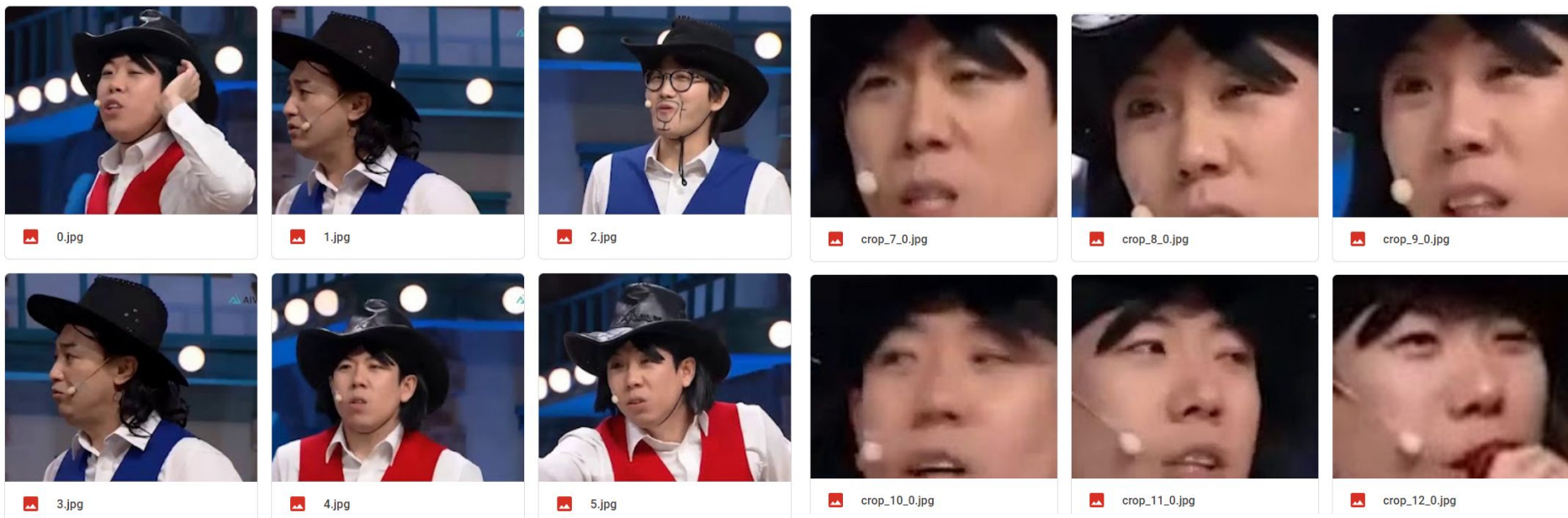
Facial landmark 좌표추출



Facial landmark을 이용한 face crop

총 데이터 구성 확인

■싸이코러스 : 400여장



싸이코러스 이미지

총 데이터 구성 확인

■코빅엔터 : 1700여장



코빅엔터 이미지

총 데이터 구성 확인

■ 결혼해두목 : 2600여장



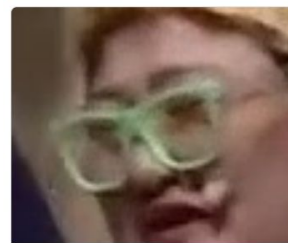
0.jpg



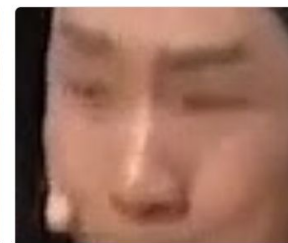
1.jpg



2.jpg



crop_37_0.jpg



crop_38_0.jpg



crop_39_0.jpg



3.jpg



4.jpg



5.jpg



crop_39_1.jpg



crop_40_0.jpg



crop_41_0.jpg

결혼해두목 이미지

총 데이터 구성 확인

■두분사망토론 : 1900여장



두분사망토론 이미지

총 데이터 : 5000여장

전처리 4 : 상단 코너 데이터 + 고정 출연진의 얼굴데이터

■위에서 전처리한 데이터를 이용하여 다음과 같이 데이터를 구성했다

- ✓ 사이코러스 : 551장
- ✓ 코빅엔터 : 1925장
- ✓ 두분사망토론 : 2125장
- ✓ 결혼해두목 : 2700장

전처리 2을 사용한 모델 학습(로고, grayscale)

■모델링 파라미터

```
Model: "sequential"
-----
Layer (type)                Output Shape              Param #
-----
conv2d (Conv2D)              (None, 480, 854, 32)      320
max_pooling2d (MaxPooling2D) (None, 160, 285, 32)      0
conv2d_1 (Conv2D)            (None, 160, 285, 32)      9248
max_pooling2d_1 (MaxPooling2D) (None, 53, 95, 32)        0
flatten (Flatten)            (None, 161120)            0
dense (Dense)                (None, 16)                2577936
dense_1 (Dense)              (None, 4)                 68
-----
Total params: 2,587,572
Trainable params: 2,587,572
Non-trainable params: 0
-----
```

■train

```
ving model to /content/drive/MyDr
loss: 0.2006 - accuracy: 0.9458
```

■val

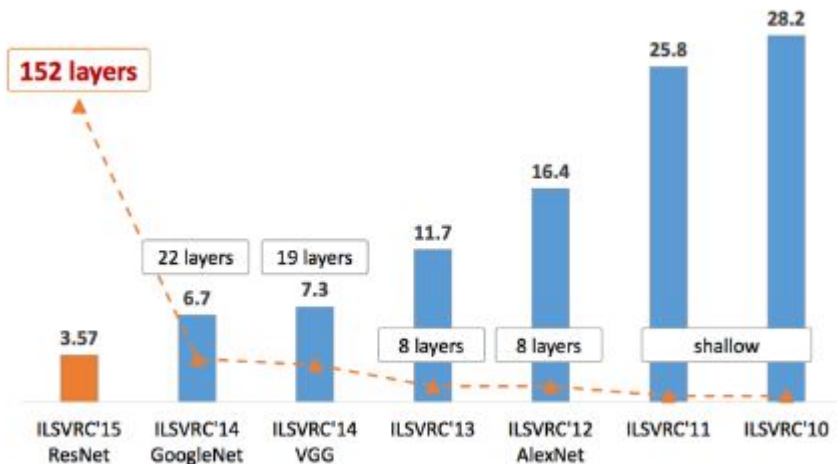
```
val_loss: 0.0638 - val_accuracy: 0.9919
```

```
Epoch 2/100
15/15 [=====] - ETA: 0s - loss: 0.2006 - accuracy: 0.9458
Epoch 2: val_loss improved from 0.47005 to 0.06377, saving model to /content/drive/MyDrive/AIVLE_3rd_Team/model
15/15 [=====] - 25s 2s/step - loss: 0.2006 - accuracy: 0.9458 - val_loss: 0.0638 - val_accuracy: 0.9919
Epoch 3/100
```

모델 설명

최종 모델 ResNet 선정 이유

✓ ResNet 성능



ResNet 이전까지 많이 사용되던 GooLeNet 보다 더 간단한 CNN 구조와 3.57 % 낮은 오차율

✓ 모델 요약(구조 등)

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

총 50개의 layer를 가지고 있으며, layer마다 다른 residual block 형태가 반복되어 학습되는 과정

전처리 4를 사용한 모델 파라미터

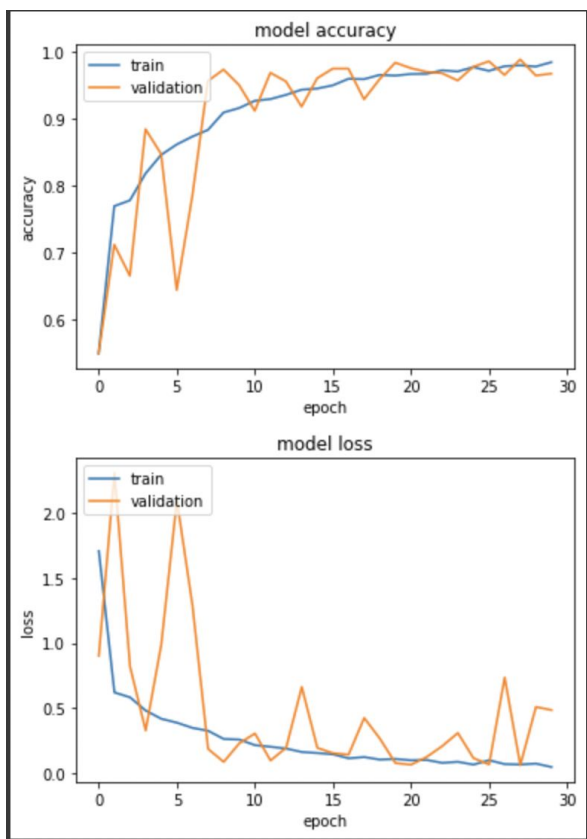
▪ResNet50 params

- ✓ Total params : 32,009,988
- ✓ Input_size : (256, 256, 3)
- ✓ loss : categorical_crossentropy
- ✓ Optimizer : Adam
- ✓ batch_size : 4
- ✓ train : val = 8 : 2
- ✓ epoch = 30
- ✓ learning_rate = 0.001

전처리 4를 사용한 모델 결과

■ResNet50 model result

```
Epoch 28: val_loss improved from 0.06254 to 0.05821, saving model to /content/drive/MyDrive/AIVLE_3rd/model/팀_미니프로젝트3차_a026198_신동명.h5  
1461/1461 [=====] - 105s 72ms/step - loss: 0.0638 - accuracy: 0.9803 - val_loss: 0.0582 - val_accuracy: 0.9890
```



✓ model accuracy graph

✓ model loss graph

모델 검증

- ✓ 모델은 과연 어떤 특징을 보고 예측을 할까?
- ✓ 로고와 사람의 얼굴중 어떤것을 보고 예측을 할까?
- ✓ Grad-Cam을 사용해 보자 !!

Grad-Cam이란?

- ✓ CNN의 내부를 열어볼 수 있는 아주 유용한 도구!
- ✓ 뉴럴 네트워크가 이미지의 어떤 부분을 보고 특정 레이블로 판단을 내리는지 확인할 수 있다.

original_image



Grad-CAM



전처리 4번의 Grad-Cam 확인

✓ 오프닝



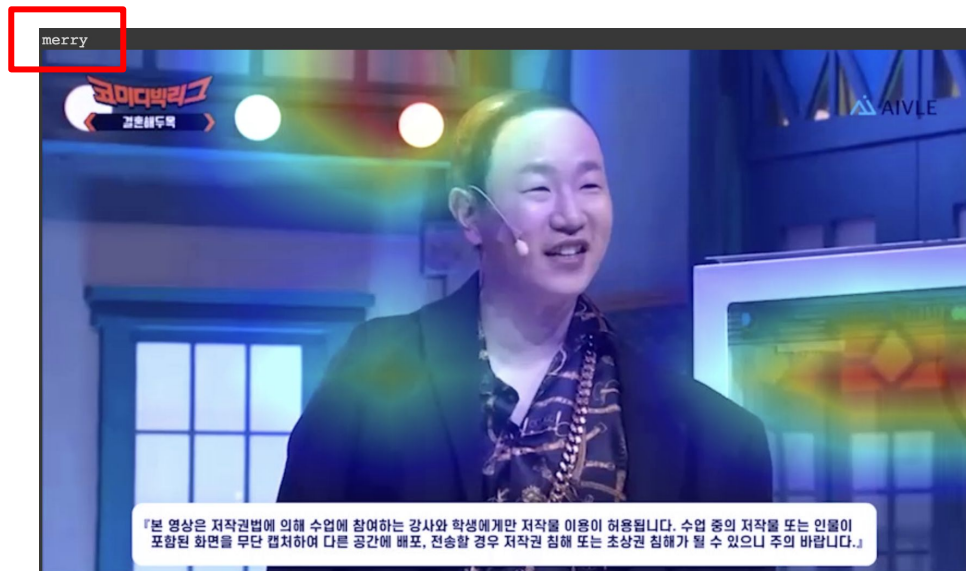
전처리 4번의 Grad-Cam 확인

✓ 코빅엔터



전처리 4번의 Grad-Cam 확인

✓ 결혼해두목



전처리 4번의 Grad-Cam 확인

✓ 두분사망토론



전처리 4번의 Grad-Cam 확인

✓ 결과발표



모델 적용 결과

■ 모델로 TEST 파일을 분류해낸 결과에 대한 설명

✓ 테스트 데이터 시간대별 코너 분포

오프닝 : 02:08

코빅엔터 : 02:08 ~ 17:18

사이코러스 : 17:18 ~ 25:55

결혼해두목 : 25:55 ~ 39:17

두분사망토론 : 39:17 ~ 49:09

✓ 로고 부분만 잘라서 적용(gray-scale)한게 가장 정확, 로고가 없는 부분은 학습이 되어있지 않아 맞추지 못함

1030	enter
1031	enter
1032	enter
1033	enter
1034	enter
1035	enter
1036	enter
1037	enter
1038	enter
1039	enter
1040	enter
1041	enter
1042	enter
1043	chorus
1044	chorus
1045	chorus
1046	chorus
1047	chorus
1048	chorus
1049	chorus
1050	chorus

sec	
1551	chorus
1552	chorus
1553	chorus
1554	chorus
1555	chorus
1556	enter
1557	enter
1558	merry
1559	merry
1560	merry
1561	merry
1562	merry
1563	merry
1564	merry
1565	merry
1566	merry
1567	merry

sec	
2351	merry
2352	merry
2353	merry
2354	merry
2355	merry
2356	merry
2357	merry
2358	merry
2359	enter
2360	enter
2361	two
2362	two
2363	two
2364	two
2365	two
2366	two
2367	two
2368	two
2369	two
2370	two
2371	two
2372	two
2373	two
2374	two

kt

 AIVLE

 AIVLE
Let's make it possible