

뮤직뱅크 AI 직캠 제작 및 깊이 추정을 통한 직캠 인물 검출 성능 향상

(A)I-DLE | 송준호, 라윤경, 박범준, 최한비

목차

1 연구 배경

⋮

1. AI 직캠이란?
2. 객체 탐지
3. 데이터셋 소개

2 문제 제기

⋮

1. 데이터 정제 과정의 과제
2. 기존 방식의 해결방안

3 해결 방안

⋮

1. 깊이 추정
2. 깊이 추정을 통한 데이터 정제
3. 데이터 정제 순서도
4. 추가 연구: 데이터 증강

4 연구 결과

⋮

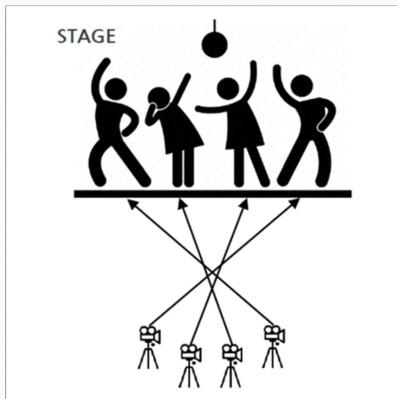
1. 평가 지표: mAP
2. 성능 비교
3. 결론

AI 직캠이란?

멤버 별 개인 직캠 서비스(참고: KBS Kpop 유튜브 채널)

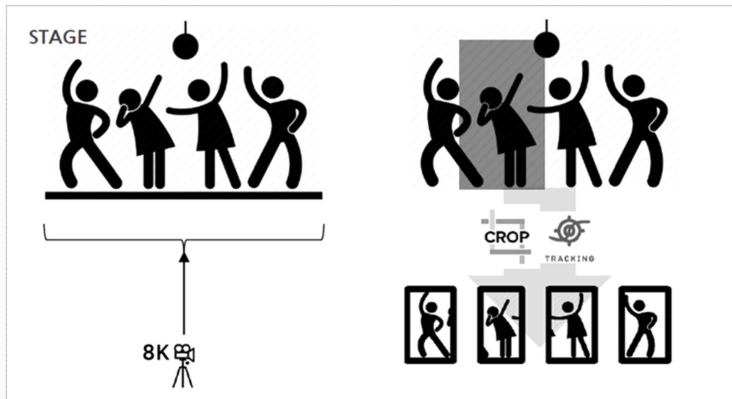
	[K-Fancam] 아이브 장원영 직캠 '해야' (HEYA)'(IVE WONYOUNG Fancam) @뮤직뱅크 (Music Bank) 240503 KBS Kpop • 조회수 7.5만회 • 11일 전
	[K-Fancam] 아이브 안유진 직캠 '해야' (HEYA)'(IVE YUJIN Fancam) @뮤직뱅크(Music Bank) 240503 KBS Kpop • 조회수 8.9만회 • 11일 전
	[K-Fancam] 아이브 레이 직캠 '해야' (HEYA)'(IVE REI Fancam) @뮤직뱅크(Music Bank) 240503 KBS Kpop • 조회수 2.4만회 • 11일 전
	[K-Fancam] 아이브 리즈 직캠 '해야' (HEYA) (IVE LIZ Fancam) @뮤직뱅크(Music Bank) 240503 KBS Kpop • 조회수 3.7만회 • 11일 전

AI 직캠이란?



기존 방식

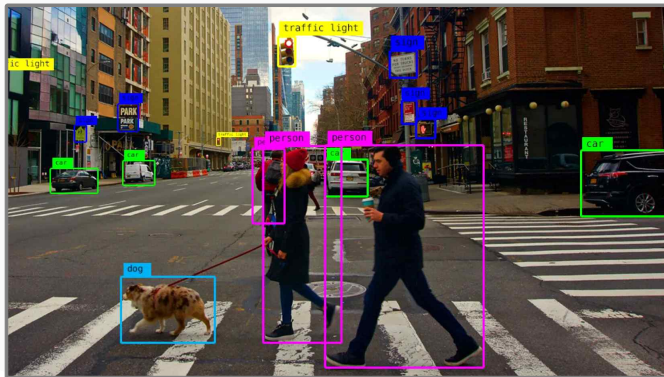
아티스트 1인당 카메라 1대와
카메라맨 1명



AI 직캠

객체 탐지 기반의 자동 화면 리프레이밍(Re-Framing) 기술
→ 촬영 비용 감소 및 무대 공간 절약

객체 탐지(Object Detection)

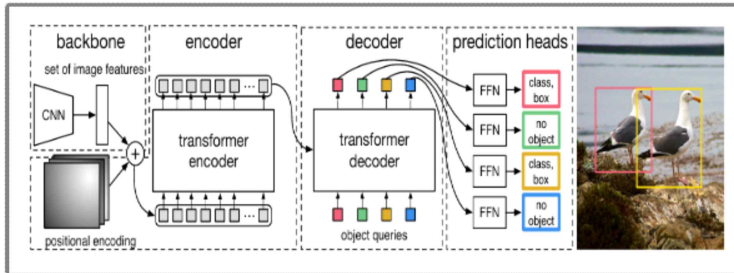


객체 탐지:

복수의 객체를 분류하고 위치를 파악하는 작업을 수행하는 컴퓨터 비전 기술의 분야

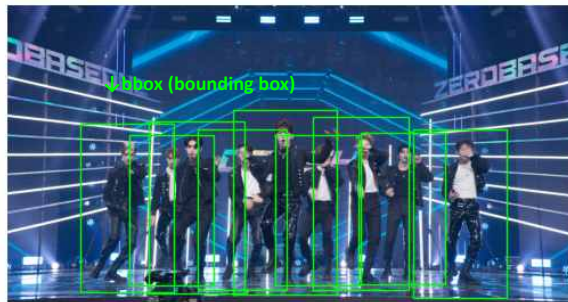
객체의 위치를 찾는 것이기에
해당 객체의 경계 상자(Bounding Box)를 표시하여 위치를 나타낸다

CNN 기반 모델이 연구되고, 최근 Transformer를 결합한 DETR 기반의 객체 검출 기법들이 등장하였다.



데이터셋 소개

2만장의 뮤직뱅크 무대 이미지 데이터와 멤버들의 위치 좌표



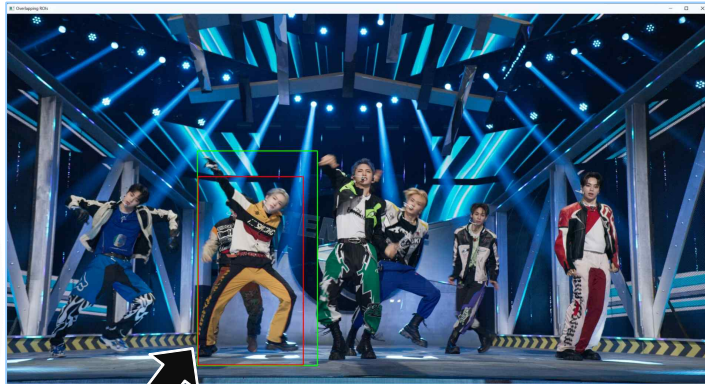
멤버별 경계 상자 시각화

```
0 건욱 0.6995516419410706 0.6823410599320023 0.14982497692108154 0.5100722701461227
0 규빈 0.2900218889117241 0.6871436931468823 0.14861376583576202 0.5059706793891059
0 김지웅 0.33417803794145584 0.6773388332790798 0.1702633947134018 0.5794889096860533
0 김태래 0.548188179731369 0.6871998539677373 0.14835530519485474 0.5049109564887153
0 리키 0.2121119424700737 0.6796527438693576 0.1661069542169571 0.56572344744647
0 매투 0.6247342824935913 0.6635656427454065 0.1701352596282959 0.5792584172001591
0 장하오 0.49043630063533783 0.6609837143509476 0.18269851803779602 0.6219896387170862
0 한빈 0.41531629860401154 0.6847868460196036 0.15773966908454895 0.5369760583948207
0 한유진 0.800221860408783 0.6974374135335286 0.16633498668670654 0.5662457784016927
```

데이터 정제 과정의 과제

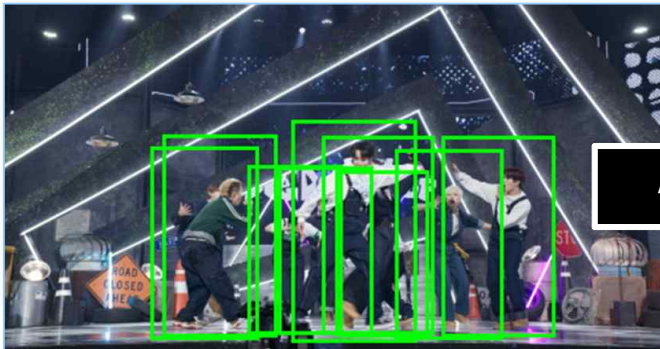


- 객체가 무대 밖을 벗어나는 경우



- 객체가 겹치는 경우

기존 방식의 해결방안



데이터 정제



겹치는 데이터를 제거하여 객체의 정보를 제거하는 방법을 사용

겹치는 인물의 정보가 유의미함에도 제거되어
정보의 손실 발생

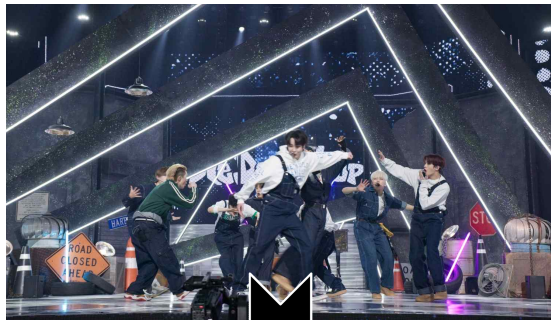
깊이 추정(Depth Estimation)

깊이 추정:

2D 이미지로부터 3D 깊이 정보를 추정하는 기술

- Monocular Depth Estimation: 단일 이미지 사용
- Stereo Depth Estimation: 두 개의 이미지 사용

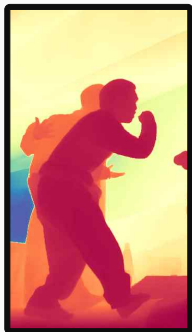
Stable Diffusion 기반 모델 Marigold 사용



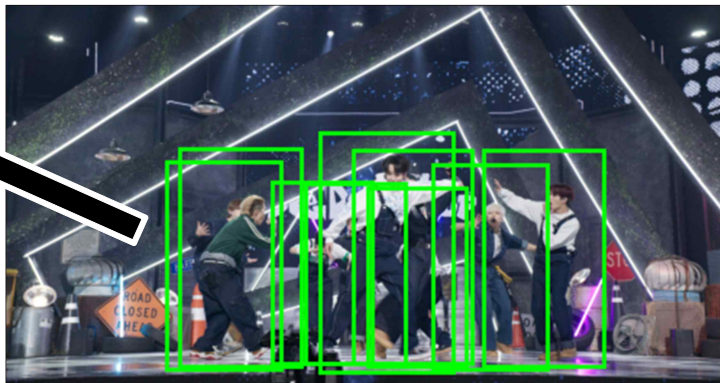
깊이 추정을 통한 데이터 정제



0.2689686



0.31226084



앞 객체의 경계 상자일수록 경계 상자가 더 큼
-> 뒷 배경이 더 많이 포함됨

뒤에 있는 객체일수록 픽셀 값은 크게 됨

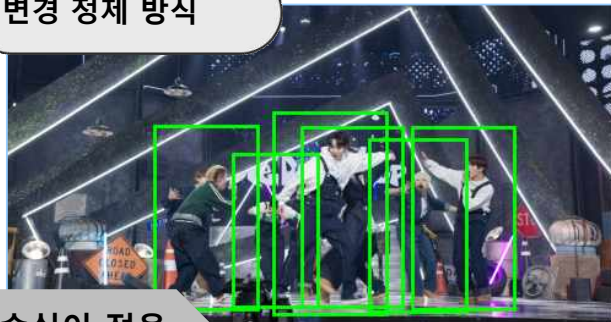
즉, 뒷 배경이 더 많이 포함될수록 평균 픽셀값이 큼
픽셀의 평균값이 가장 큰 경계 상자가 가장 앞의 경계 상자

깊이 추정을 통한 데이터 정제

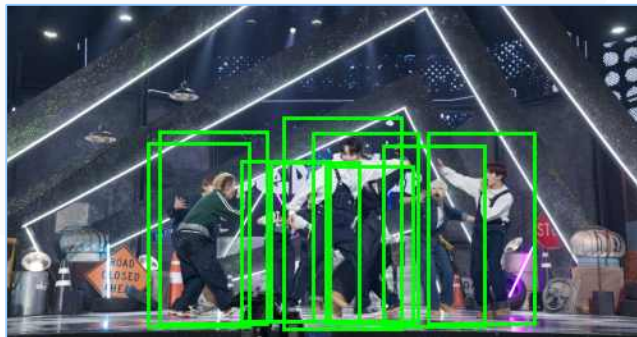
기존 정제 방식



변경 정제 방식

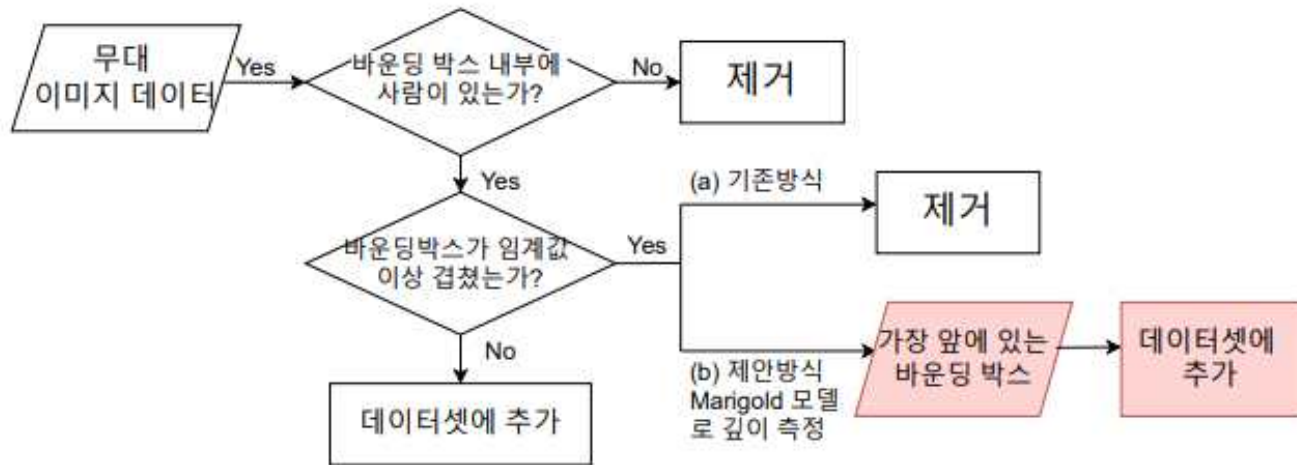


데이터 손실이 적음



• 원본 데이터

데이터 정제 순서도



추가 연구: 데이터 증강

Base Augmentations



데이터 증강(Data Augmentations)

훈련 데이터의 다양성을 늘려서, 모델이 더 넓은 조건에서 잘 작동할 수 있게 만드는 방법

이미지 인식이나 분류 작업에서 모델의 일반화 능력이 향상되고, 과적합을 줄일 수 있음

Crop, Resize, Flip, Grayscale..

추가 연구: 데이터 증강

데이터 증강 - Grayscale

- 이미지를 회색조로 바꾸는 데이터 증강 방법
- 조명색이나 옷의 색 등은 중요하지 않음
- 객체가 사람인지 판별이 목적
- 기존 데이터 증강 방식에 50% 확률로 Grayscale 을 적용한 데이터 증강을 추가하여 모델의 일반화 성능 향상을 기대



평가지표: mAP(mean Average Precision)

mAP: 객체 검출 모델의 성능 평가 지표

정밀도(Precision)와 재현율(Recall)을 결합한 지표

- 정밀도: 검출된 객체 중 실제 정답 비율
- 재현율: 실제 정답 중 검출된 비율

계산 방법

- 각 클래스별 Precision-Recall 커브 아래 면적(AP) 계산
- 모든 클래스의 AP 평균 → mAP

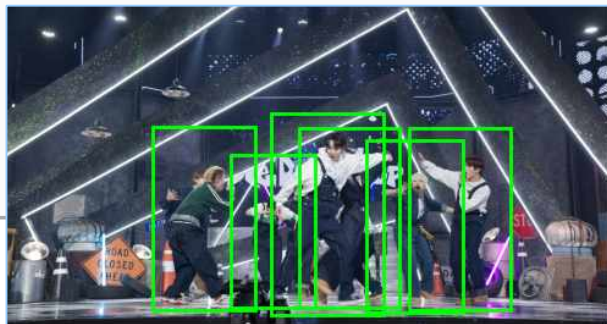
ex) mAP 0.5 (IoU 0.5 기준), mAP 0.5:0.95 (IoU 0.5~0.95 평균)

→ mAP가 높을수록 객체 검출 성능이 우수함

성능 비교

	epochs	mAP	mAP_50	mAP_75
baseline	12	0.472	0.689	0.581
diffusion	12	0.512	0.744	0.621
diffusion + grayscale	12	0.519	0.760	0.630

결론



Diffusion 기반의 **깊이 추정 모델**을 이용하여 데이터 정제 과정에서 **가장 앞에 위치한 인물 데이터를 보존**하는 새로운 방법을 제안한다.

인물 간의 깊이를 추정하여 **가장 앞의 인물 정보를 보존**하여, 무대 위 복잡한 동선 및 겹침 상황에서 전면의 인물 검출이 가능하다.

모델의 일반화 성능 향상을 위한 Grayscale 증강 방법을 제안한다.

시연 영상

TWS
NMIXX

Thank you

Q&A

Diffusion 모델 기반 깊이 추정을 통한
K-POP 직캠 인물 검출 성능 향상