

# Semi-Supervised Object Detection with Sparsely Annotated Dataset

## 0. Abstract

컨볼루션 신경망에 기반한 물체 감지기 훈련에서 훈련을 위한 효과적인 양성 예제의 선택은 중요한 요소입니다. 그러나 이미지에 스파스 주석을 가진 앵커 기반 검출기를 훈련할 때 효과적인 긍정적인 예를 찾기 위한 노력은 훈련 성과를 방해할 수 있습니다. 주어진 IoU 하에서 긍정적인 예를 수집하기 위해 ground truth bounding box에 대한 앵커 기반 훈련을 사용하는 경우, 종종 현재 교육 클래스에 다른 클래스의 개체를 포함시킬 수 있으며, 또는 훈련해야 하는 물체는 부정적인 예로만 샘플링될 수 있습니다. 우리는 이 문제를 해결하기 위해 두 가지 접근방식을 사용했습니다. 1) 앵커리스 물체 감지기와 2) 단일 물체 추적기를 사용한 준감독 학습 기반 물체 감지. 제안된 기법은 연속 프레임을 위해 시간 영역의 앵커로서 스파스 주석을 단 경계 상자를 사용하여 단일 객체 추적을 수행합니다. 추적 결과에서 트레이닝용 영상에 대한 조밀한 주석이 자동으로 생성되어 객체 검출기(Detector) 트레이닝에 사용되었습니다. 제안된 단일 객체 추적 기반 준감독 학습을 Epic-Kitchens 데이터 세트에 적용했습니다. 그 결과, Unseen 섹션에서 2위를 차지했고, IoU > 0.5 평가에서 Epic-Kitchens 2020 객체 감지 과제의 Seen 섹션에서 1위를 차지했습니다.

## 1. Introduction

CNN(Convolution Neural Networks)의 빠른 발전으로 CNN을 이용한 물체 인식 네트워크의 성능도 크게 향상되었습니다. 객체 감지 네트워크의 성능이 향상됨에 따라 PASCAL VOC와 같이 복잡도가 낮은 데이터셋에서 시작하여 MS-COCO와 같이 복잡도가 높은 데이터셋으로 개발되었습니다. 객체 감지 데이터셋 중 비교적 최근에 출시된 Epic-Kitchens 데이터셋은 다른 객체 감지 데이터셋과 다른 다음과 같은 특성을 가지고 있습니다.

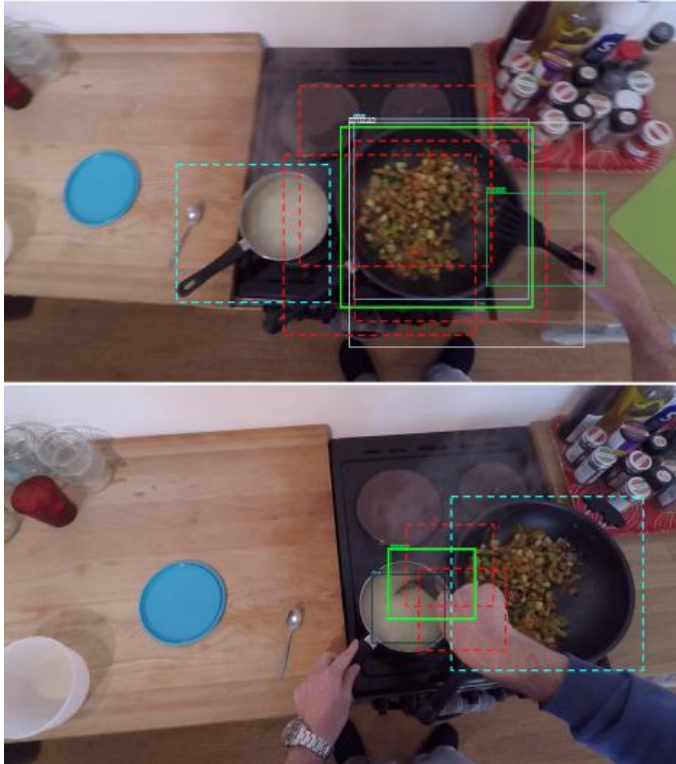


그림 1. 스파스 주석을 단 데이터 세트에 대한 앵커 기반 검출기 훈련의 예. 녹색 실선은 각 교육 영상의 라벨 정보를 나타내며, 빨간색 점선은 긍정적인 예시의 예입니다. 밝은 파란색 점선은 다른 교육 영상(상단)의 라벨에 포함되지만 훈련할 지정된 영상(하단)에는 라벨이 부착되지 않은 개체를 나타냅니다. 따라서 Epic-Kitchens 객체 감지 데이터셋에서는 앵커 기반 검출기(Anchor-Based Detector)를 교육할 때 학습해야 할 대상이지만 라벨 정보가 누락되어 교육 성능이 저하됩니다.

- 교육용 검출기(Detector) 영상이 원본 비디오에서 수집되고 해당 프레임 시퀀스가 제공됩니다.
- 교육용 이미지에서, 훈련 가능한 물체 중 일부에만 주석이 드문드문 붙어 있습니다.
- 교육 데이터 세트에 포함된 개체의 모양 분포에 따라 몇 개의 스냅샷 클래스와 여러 개의 스냅샷 클래스 간의 주석 양 차이가 큼니다.

위에서 설명한 것처럼 Epic-Kitchens의 객체 감지 주석은 기존 데이터 집합과 다른 방식으로 제공되며, 기존 객체 감지 모델을 그대로 트레이닝하는 방법을 적용하기 어렵다는 특징이 있습니다. 일반적으로 RPN의 구조로 후보 물체의 객관성을 단련하는 앵커나 검출기에 기초한 긍정적인 예를 훈련하는 검출기의 경우 효과적인 훈련을 위해 지상 진실 경계 상자가 있는 교차로(IoU)를 고려하여 배치 샘플링이 수행됩니다. 그러나 닛 기반 하드 예제 마이닝이 주석이 거의 없는 훈련 이미지에 수행될 경우, 지상 진실 경계 상자 근처의 물체 분포로 인해 훈련의 효율성이 저하됩니다. 그림 1은 스파스 주석 영상이 있는 앵커 기반 객체 감지기를 트레이닝할 때 좋은 긍정적인 예를 선택하려고 할 때의 부정적 효과를 보여줍니다.

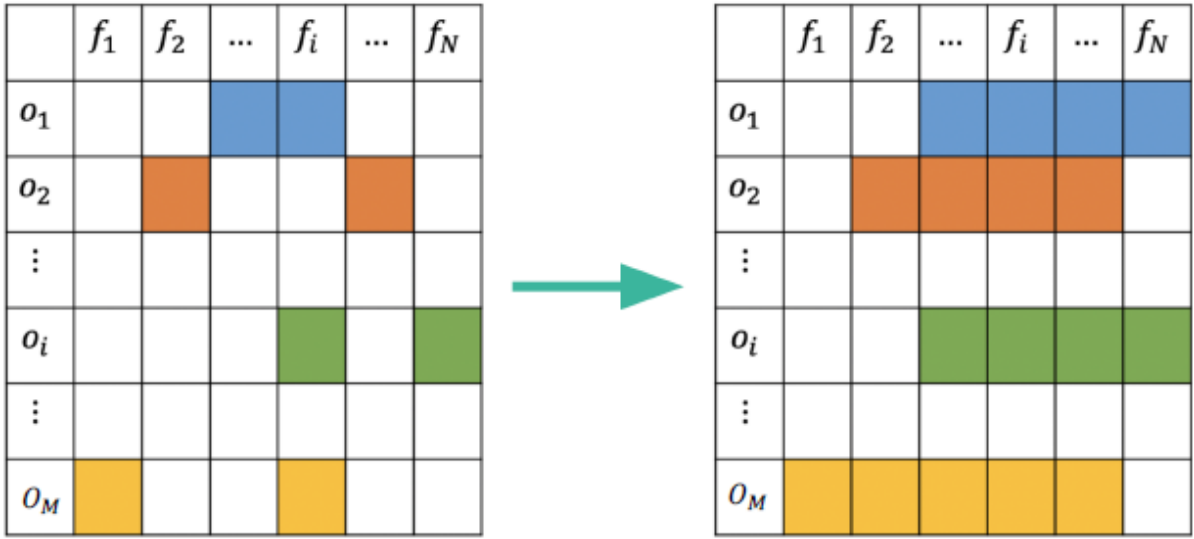


그림 2. 제안된 지도 학습의 최종 목표  $f$ 는  $N$ 개의 프레임으로 구성된 액션 클립을 의미하며  $o$ 는 액션 클립에 존재하는 총  $M$ 개의 물체를 의미한다. 우리는 액션 클립에 있는 모든 학습 가능한 물체에 대한 조밀한 라벨을 얻기 위해 양방향 추적을 통한 준감독 학습을 수행했습니다.

이 문제를 해결하기 위해 두 가지 접근 방식을 통해 물체 감지기를 교육했습니다. 첫째, 앵커 기반 검출 모델 대신 경계 상자 추정을 위해 다른 파라미터화를 사용하는 검출기가 사용되었습니다. 우리는 FCOS (Fully Convolution One-Stage Object Detection) 네트워크를 활용하여 앵커 기반 샘플링에 의해 영향을 받는 스파스 주석의 트레이닝 이미지의 영향을 최소화할 수 있었습니다. 두 번째는 Epic-Kitchens 데이터 세트의 기능을 활용하는 것입니다. 객체 추적 기술은 각 교육 이미지에서 감지 가능한 모든 물체를 세밀하게 감독하는 데 사용됩니다. 이를 위해 특정 프레임에 존재하는 경계 상자 라벨은 시간 영역의 초기 경계 상자로 설정되고 단일 개체 추적기의 입력으로 사용됩니다. 그 후, 트래커의 임계값을 초과하는 예측 출력은 폐소 주석으로 간주되었고, 모든 학습 가능한 물체에 대한 라벨이 훈련을 위해 영상에 제공되었습니다. 그림 2는 준감독 학습의 제안된 목표를 도식적으로 보여줍니다.

제안된 접근 방식을 통해 우리는 Epic-Kitchens 객체 감지 데이터 세트를 통해 객체 감지 네트워크를 효과적으로 교육할 수 있었습니다. 그 후, 트레이닝 받은 비균질 백본이 있는 FCOS 모델에 대해 합동 NMS 기반 앙상블이 수행되었습니다. 그 결과, Epic-Kitchens 2020의  $\text{IoU} > 0.5$ 에 따른 Seen 평가 세트에서 1위를 차지했고 Unseen 평가 세트에서 2위를 차지했습니다.

## 2. Related Work

**물체 감지.** CNN 기반 물체 탐지 모델은 크게 1단계 모델과 2단계 모델로 나뉩니다. 1단계 모델에서 물체의 등급과 위치를 예측하는 과정은 하나의 구조로 이루어지며, YOLO, SSD, RetinaNet 등이 그 예입니다. 기능 피라미드 네트워크 (FPN)와 같은 추가 구조는 헤드 구조의 백본에서 얻은 출력 특징을 효율적으로 처리하기 위해 종종 사용됩니다. 일반적으로 분류와 회귀는 하나의 구조

로 이루어지기 때문에 회귀 정확도가 2단계 모델보다 낮은 것으로 알려져 있습니다. 2단계 모델의 경우, 물체 위치에 대한 사전 지식은 검출기의 하위 네트워크인 RPN(지역 제안 네트워크)에서 추정합니다. RPN은 클래스에 구애받지 않는 서브넷으로 객관성을 결정하고 이후의 헤드 구조를 통해 클래스 인식 탐지를 수행합니다. R-CNN, R-FCN, 캐스케이드 R-CNN, 캐스케이드 RPN 등이 다양한 헤드 구조를 대표하며 상대적으로 높은 회귀 정확도를 갖는 것으로 알려져 있습니다. 1단계 및 2단계 모델의 철학을 결합하는 정밀 검출기(RefineDet)와 같은 모델도 제안되었으며 FCOS와 같은 구조적 장점보다는 경계 박스 회귀 분석을 위한 다른 매개변수화를 활용하는 검출기(Detector) 또한 제안되었습니다.

**물체 감지를 위한 준감독 학습.** 준감독 학습을 사용한 물체 감지는 학습하기에 충분한 수의 주석을 수동으로 획득하기 어렵거나 상대적으로 많은 수의 라벨 미표시 데이터에서 유사 라벨을 획득해야 하는 상황에서 사용됩니다. [18] 연속 프레임에서 양호한 유사 라벨을 얻기 위해 사전 훈련된 물체 감지기와 강력한 추적기를 사용하여 유사 라벨을 평가하고 재훈련하기 위한 반복 프레임워크를 제안했습니다. [19]에서는 주석이 드문 대규모 데이터셋에 대한 양호한 의사 라벨을 얻기 위해 부품 인식 샘플링 및 ROI 제안을 활용하여 Open Image Dataset V4에서 검출 성능을 개선할 수 있었습니다. [5]에서 MS-COCO 데이터 세트의 라벨링되지 않은 데이터를 효율적으로 사용하기 위해, 동시 매트릭스 분석을 사용하여 라벨링된 데이터 세트의 이전 정보를 사용하여 양호한 의사 라벨을 생성했습니다. 제안된 단일 객체 추적기 기반 준감독 학습은 추적기를 사용한다는 점에서 [18]와 유사하지만 기존의 희박한 주석 정보를 사용하여 특정 영상에 대한 조밀한 주석 정보를 얻는 차이는 있습니다. 동시에 객체 감지기는 추적을 위한 초기 입력에 사용되지 않기 때문에 훈련은 반복 훈련 시나리오로 적용되지 않습니다.

**단일 객체 시각적 추적.** 우리는 단일 객체 추적 네트워크를 사용하여 스파스 주석을 단 데이터셋에 대한 의사 라벨을 생성했습니다. 단일 객체 추적[1, 13, 27, 3, 12]에서 삼 네트워크 기반 시각적 추적기는 다양한 데이터셋에서 균형 잡힌 정확도와 속도를 보여줍니다. 삼 네트워크 기반 추적기는 기본적으로 대상 이미지에 대한 CNN 기능과 추적에 대한 입력 이미지의 유사성을 가지고 트레이닝됩니다. 우리는 SiamMask를 단일 객체 추적기로 사용했는데, 추적 대상과 기능의 유사성과 함께 상자 및 마스크 정보를 사용합니다.

### 3. Fully Convolutional One-Stage Object Detection (FCOS)

우리는 FCOS 모델을 사용해 검출기 훈련에서 앵커를 사용한 양호한 양성 예를 선택하기 위한 계산 프로세스를 제외했습니다. FCOS는 경계 상자의 회귀에 대한 파라미터를 다르게 정의했으며 앵커 프리 검출기를 제시했습니다. FCOS의 손실 기능은 다음과 같이 정의됩니다.

$$L(a_p, m_p) = \frac{1}{N} \sum_{\{u,v\} \in p} L_{cls}(a_p, c_p) + \frac{1}{N} \sum_{\{u,v\} \in p} [c_p > 0] L_{reg}(m_p, \hat{m}_p), \quad (1)$$

여기서  $p$ 는 특성 맵의 위치  $(u, v)$ 를 나타내며,  $a_p$ 은 클래스 추정을 위한 예측 벡터를 나타냅니다.  $c_p$ 는 입력 예제의 클래스 라벨을 나타내며,  $m_p$ 는 특성 맵의 특정 공간 위치와 지상 진실 경계 상자  $m_p = (l, t, b)$ 로부터의 거리를 나타냅니다. 경계 상자에 대한 라벨  $B = (x_0, y_0, x_1, y_1, c)$ 가 주어진 경우 FCOS는 경계 상자를 찾기 위해  $l = u = u = v = y_0, r = u = v = v = y_1$ 에 의해 매개변수화합니다. 우리는 FCOS를 위한 백본 네트워크로 사전 교육받은 ResNet, ResNeXt, HRNet을 사용했습니다.

#### 4. Semi-Supervised Learning with Single Object Tracker

Epic-Kitchens 데이터 집합은 특정 개체에 대한 경계 상자 레이블과 개체가 나타나는 작업 클립에 대한 시퀀스 프레임을 동시에 제공합니다. 개체의 경계 상자는 모든 프레임이 조밀하게 지정된 것이 아니라 작업 순서에서 드문드문 나타납니다. 그림 2에서는 단일 객체 추적기를 사용하여 자동화된 절차를 통해 목표를 달성했습니다. 다양한 단일 객체 추적기 중에서 SiamMask가 사용되었으며, 이는 추적 정확도와 속도를 위한 균형 잡힌 성능을 보여줍니다. DAVIS 데이터 집합에서 학습한 SiamMask 모델을 사용하여 각 경계 상자를 단일 객체의 초기 값으로 사용하여 양방향 추적을 수행했습니다. 하나의 작업 클립 입력에 대해 SiamMask를 사용한 전방 추적에 대한 자세한 내용은 알고리즘 1에 설명되어 있습니다. 알고리즘 1은 역추적을 통해 양방향 추적을 완료하는 것과 동일한 방식으로 사용됩니다. 그림 3은 단일 객체에 대한 알고리즘 1에 따른 추적의 시작과 끝의 예를 보여주고, 그림 4는 객체 추적 후 생성된 의사 라벨이 있는 교육 이미지를 보여줍니다.

---

##### Algorithm 1: Forward tracking

---

Input: Action clip ( $A$ ), pretrained tracking model ( $T$ ),  
a set of bbox for initial input ( $BB$ ), threshold of  
tracking score ( $\rho_1$ ), threshold of IoU between two  
pair of tracked bbox ( $\rho_2$ )  
Output:  $BB$  in  $Q$  from  $T$   
Initialize an empty queue  $Q$   
**while** bbox  $b_{c,i}$  with class  $c$  at  $i$ -th frame available  
from  $BB$  **do**  
    Get a list of frames  $FF$  in forward from  $i$ -th frame  
    in  $A$ ;  
    Initialize  $T$  with  $b_i$  from  $A$ ;  
    Initialize a variable  $prev_s$  with a size of  $b_{c,i}$  to store  
    a size of object from  $T$  at previous frame;  
    **while** each frame in  $FF$  **do**  
        Get a bbox  $b_{c,k}$  from  $T$  at  $k$ -th frame;  
         $crnt_s :=$  a size of  $b_{c,k}$ ;  
        **if**  $IoU(prev_s, crnt_s) \geq \rho_1$  **then**  
             $prev_s := crnt_s$ ;  
            Add  $b_{c,k}$  to  $Q$ ;  
        **else**  
            **break**;  
        **end**  
    **end**  
**end**

---



그림 3. 알고리즘 1의 결과 예제입니다. 파란색 상자로 표시된 프레임은 추적이 시작된 이후 동일한 개체로 추적된 프레임이며, 빨간색 상자로 표시된 프레임은 알고리즘 1의 종료 조건으로 인해 추적이 종료된 프레임입니다.



그림 4. 추적 후 교육 영상의 변경 사항. 세미 감독 학습에 사용할 최종 주석의 예가 빨간색 점선으로 표시된 교육 영상에 표시됩니다.

표 1. 각 디텍터의 교육 세부 정보. 'lr'는 학습률을 나타냅니다. 학습 일정은 (scheduler, drop rate) [drop epoch1:drop epoch2:max epoch]로 표시됩니다.

Detector	Backbone	Optimizer (lr)	Learning schedule	Warmup(iter, ratio)
Faster R-CNN	ResNet-101	SGD (0.02)	(step, 0.1)[8:11:24]	linear(500, 1/3)
Cascade R-CNN	HRNet-V2P	SGD (0.02)	(step, 0.1)[16:19:30]	linear(500, 1/3)
FCOS	HRNet-V2P	SGD (0.01)	(step, 0.1)[8:11:20]	constant(500, 1/3)
FCOS	ResNet-50	SGD (0.01)	(step, 0.1)[8:11:16]	constant(500, 1/3)
FCOS	ResNet-101	SGD (0.01)	(step, 0.1)[16:20:22]	constant(500, 1/3)
FCOS	ResNeXt-101	SGD (0.01)	(step, 0.1)[16:22:39]	constant(500, 1/3)

표 2. 앵커 기반 및 앵커리스 검출기의 성능 비교. 모델 이름에 '+'가 있으면 트래커 기반 준감독 학습을 사용한 평가 결과입니다. 단일 모델에서 최고의 성능과 전체 모델에서 최고의 성능을 볼드체로 표시합니다.

Detector	backbone	Seen			Unseen		
		> 0.05	> 0.5	> 0.75	> 0.05	> 0.5	> 0.75
Faster R-CNN	ResNet-101	37.54	28.64	6.92	32.83	23.16	5.55
Cascade R-CNN	HRNet-V2P	30.44	24.17	8.73	23.87	18.05	6.81
FCOS	HRNet-V2P	48.44	34.87	<b>11.02</b>	<b>43.88</b>	<b>30.68</b>	<b>9.27</b>
FCOS	ResNet-50	46.96	34.51	10.09	42.46	29.49	7.48
FCOS	ResNet-101	49.77	35.8	10.15	43.39	28.98	7.86
FCOS	ResNeXt-101	48.17	33.95	9.86	41.79	27.27	7.19
FCOS+	ResNet-101	<b>50.27</b>	<b>35.89</b>	10.57	43.14	29.82	7.76
FCOS Ensemble+	-	<b>58.27</b>	<b>44.48</b>	<b>15.36</b>	<b>55.72</b>	<b>41.12</b>	<b>12.5</b>



## 5. Epic-Kitchens Object Detection Results

**트레이닝 세부사항.** 앵커 기반 검출기로 Faster R-CNN과 Cascade R-CNN을, 앵커리스 검출기로 FCOS를 사용하여 Epic-Kitchens 객체 검출 데이터 집합의 성능을 비교했습니다. 검출기 트레이닝을 위한 백본 CNN으로 ResNet-50, ResNet-101, ResNetXt-101, ImageNet으로 사전 트레이닝된 HRNet-V2p-W32가 사용되었으며, 백본과 머리 구조의 각 조합에 대한 교육 세부 사항은 표 1에 나와 있습니다. 모든 실험은 MMDetection 라이브러리를 사용하여 수행되었습니다.

**앵커 기반 대 앵커리스 검출기** 표 2는 앵커 기반 검출기와 앵커리스 검출기의 단일 모델에서 트레이닝 성능을 보여줍니다. 표 2에 따르면, 기초적인 트레이닝 성과에서 앵커가 없는 검출기의 성능이 우수하며 안정적인 학습 성과를 보여주고 있습니다. 그림 5는 앵커 기반 검출기와 앵커리스 검출기의 훈련 중 손실 변화를 보여줍니다. 앵커가 없는 검출기는 비교적 안정적인 손실 곡선을 보여줍니다. 동시에 표 2는 다양한 백본에 따른 FCOS 모델의 성능 변화를 보여줍니다. 단일 모델의 경우, ResNet-101 백본을 활용한 FCOS 모델이 Seen 집합에서 최고의 일반화 성능을 달성하였고, HRNet 백본 모델은 Unseen 집합에서 최고의 성능을 달성하였음이 확인되었습니다.

**준감독 학습.** DAVIS 데이터 집합의 사전 트레이닝된 SiamMask 모델을 사용하여 고밀도 레이블을 생성하여 고밀도 주석 FCOS 모델을 트레이닝했습니다. 표 2는 단일 객체 추적을 기반으로 하는 준감독 학습을 사용할 때  $IoU > 0.5$ 에 따른 FCOS 모델의 일반화 성능이 지속적으로 향상되는 것을 보여줍니다.

**백본 앙상블이 균질하지 않습니다.** 우리는 훈련된 검출기에서 최고의 검출 성능을 달성하기 위해 표 2의 각 훈련된 모델에 대해 모델 앙상블을 수행했습니다. 모든 검출기에서 얻은 최대 300점의 높은 예측 점수로 경계 상자에 NMS를 적용하여 앙상블을 달성할 수 있는 공동 NMS 기법을 사용하여 앙상블을 수행했습니다. 표 2는 앙상블 조합에 따른 Seen과 Unseen 세트의 성능 변화를 보여주고 있으며, 그림 6은 Epic-Kitchen 객체 감지 챌린지 페이지에 게재된 성능을 보여준다. 드디어  $IoU > 0.5$  평가 하의 비균질 백본 앙상블을 통해 See set 및 Unseen set에서 준우승 성과를 달성할 수 있었습니다.

**시각화.** 우리는 비균질 백본 앙상블의 효과를 확인하기 위해 각 모델의 추론 결과를 시각화했습니다. 그림 7은 서로 다른 백본으로 FCOS 모델의 추론 경계 상자의 시각화를 보여줍니다. 그림 7과 같이 검출기의 구조는 동일하지만 매우 다른 유형의 추론 결과를 얻기 위해 서로 다른 백본만 사용할 수 있습니다. 이를 통해 앙상블 모델은 단일 모델에 비해 매우 큰 성능 향상을 달성할 수 있음을 확인했습니다.

## 6. Conclusion

단일 객체 추적기 기반 준감독 객체 감지를 수행하여 시퀀스 이미지에 대한 스파스 주석이 데이터 세트를 효과적으로 트레이닝했습니다. 제안된 기법의 효용을 검증하기 위해 Epic-Kitchens 객체 감지 데이터 세트를 사용했으며, 제안된 반감독 학습은 단일 모델뿐만 아니라 앙상블에서도 좋은

성과를 보였습니다. 그러나 앵커 기반 모델의 장단점에 대한 준감독 학습으로 더 면밀히 분석할 필요가 있으며, 의사 레이블을 얻기 위해 단순한 규칙 기반 엔진을 사용하는 데는 한계가 있습니다. 향후 개선을 위해서는 앵커와 RPN이 스파스 주석 데이터 훈련에 미치는 영향에 대한 정량적 분석이 필요하며, 동시에 훈련 중 추적 규칙을 개선하거나 추적 프로세스에서 얻은 결과를 활용하는 방법을 고려할 수 있습니다.

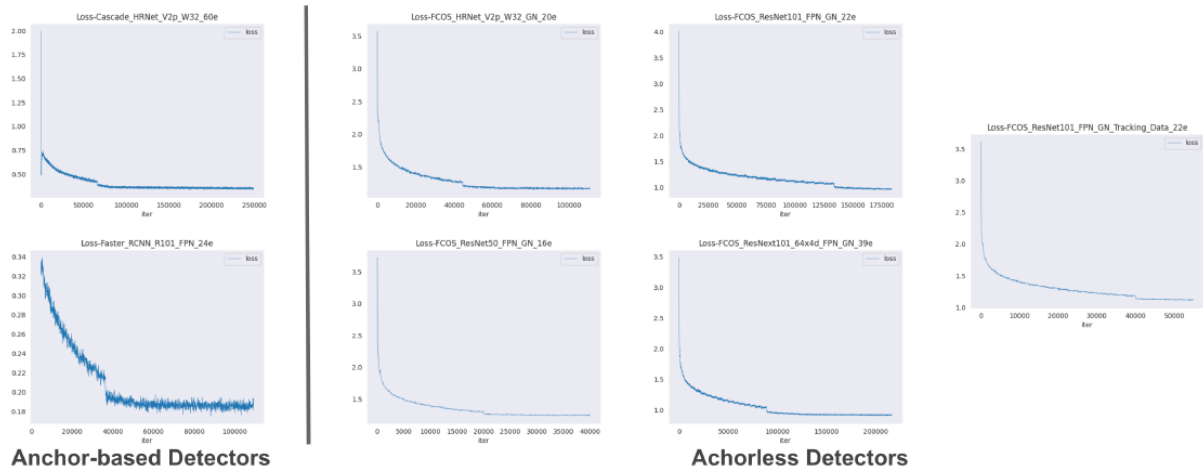


그림 5. 앵커 기반 및 앵커리스 검출기에 대한 교육 손실. 앵커리스 검출기의 손실 곡선이 상대적으로 안정적이라는 것을 보여준다.

Seen Kitchens (S1)													
#	User	Entries	Date of Last Entry	Team Name	Few Shot Classes (%)			Many Shot Classes (%)			All Classes (%)		
					IoU > 0.05 ▲	IoU > 0.5 ▲	IoU > 0.75 ▲	IoU > 0.05 ▲	IoU > 0.5 ▲	IoU > 0.75 ▲	IoU > 0.05 ▲	IoU > 0.5 ▲	IoU > 0.75 ▲
1	killerchef	51	05/30/20	hutorn	47.44 (3)	35.75 (1)	14.32 (2)	60.77 (2)	46.50 (1)	15.60 (2)	58.27 (2)	44.48 (1)	15.36 (1)
2	gongtao	35	05/19/20		49.95 (2)	32.63 (2)	6.64 (6)	60.03 (3)	44.39 (3)	9.71 (8)	58.13 (3)	42.18 (2)	9.13 (6)
3	kide	27	05/29/20	DHARI	54.98 (1)	32.40 (3)	14.55 (1)	68.74 (1)	43.88 (4)	15.38 (3)	66.15 (1)	41.72 (3)	15.23 (2)
4	gb7	69	04/01/20	FB AI	26.55 (8)	19.01 (8)	8.22 (4)	58.44 (4)	46.22 (2)	15.61 (1)	52.44 (4)	41.10 (4)	14.22 (3)
5	cvg_uni_bonn	23	05/12/20	CVG Lab Uni Bonn	39.36 (4)	26.66 (4)	7.89 (5)	53.50 (5)	41.28 (5)	12.46 (4)	50.84 (5)	38.53 (5)	11.60 (4)
Unseen Kitchens (S2)													
#	User	Entries	Date of Last Entry	Team Name	Few Shot Classes (%)			Many Shot Classes (%)			All Classes (%)		
					IoU > 0.05 ▲	IoU > 0.5 ▲	IoU > 0.75 ▲	IoU > 0.05 ▲	IoU > 0.5 ▲	IoU > 0.75 ▲	IoU > 0.05 ▲	IoU > 0.5 ▲	IoU > 0.75 ▲
1	gb7	69	04/01/20	FB AI	13.70 (8)	10.41 (8)	2.88 (7)	59.21 (2)	45.42 (1)	16.24 (1)	54.57 (4)	41.85 (1)	14.88 (1)
2	killerchef	51	05/30/20	hutorn	29.81 (3)	20.87 (4)	8.09 (1)	58.66 (3)	43.42 (2)	13.00 (3)	55.72 (2)	41.12 (2)	12.50 (3)
3	kide	27	05/29/20	DHARI	35.75 (1)	22.31 (2)	7.33 (4)	67.92 (1)	41.92 (3)	14.29 (2)	64.64 (1)	39.93 (3)	13.58 (2)
4	gongtao	35	05/19/20		35.72 (2)	25.60 (1)	7.78 (3)	56.93 (4)	41.19 (4)	8.75 (7)	54.77 (3)	39.60 (4)	8.65 (6)
5	cvg_uni_bonn	23	05/12/20	CVG Lab Uni Bonn	25.34 (4)	21.54 (3)	7.81 (2)	52.18 (5)	38.24 (5)	11.41 (4)	49.45 (5)	36.54 (5)	11.04 (4)

그림 6. Epic-Kitchen 2020 객체 감지 챌린지 평가 페이지. 빨간색 상자로 표시된 항목은 우리가 제안한 접근법에 의해 평가된 최종 성과입니다. 각 항목은 IoU > 0.5 평가에서 순위가 매겨집니다.



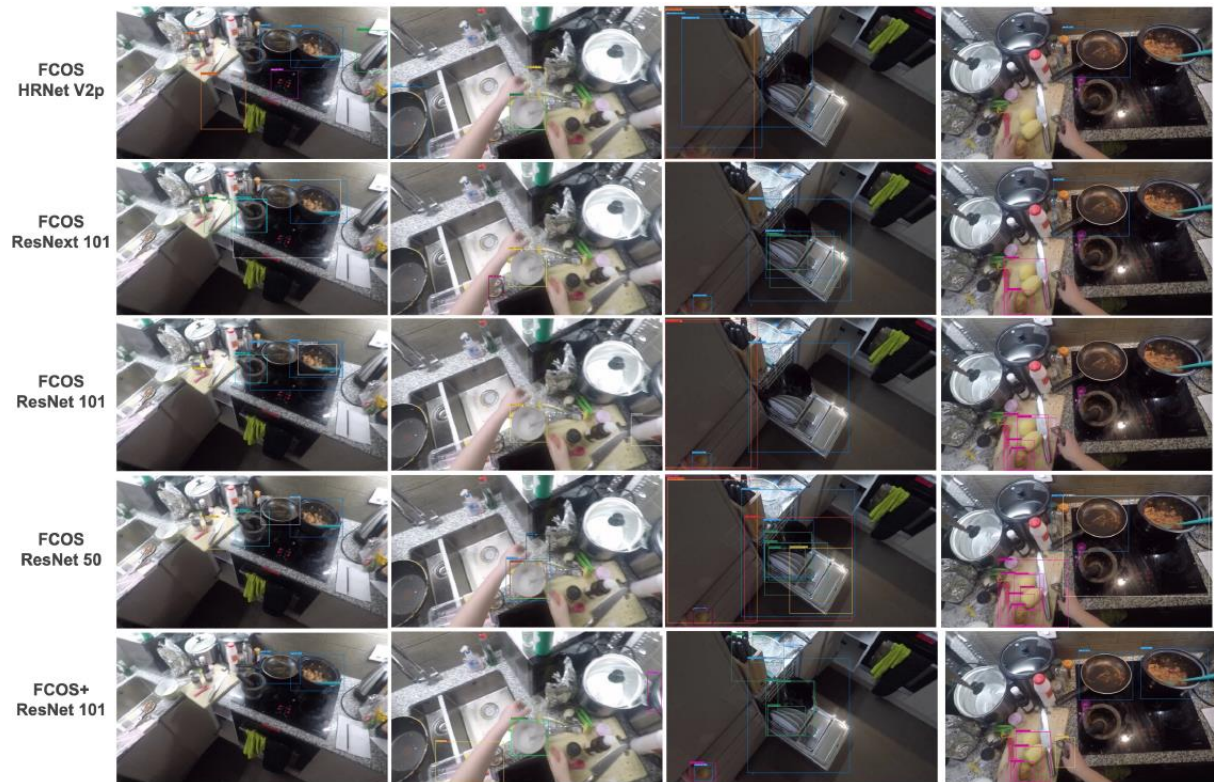


그림 7. 각 네트워크 예측에 대한 시각화. 이것은 예측 점수가 구조가 같지만 이질적인 백본을 가질 때 각 네트워크마다 다양하게 분배될 수 있음을 보여줍니다. 시각화에 대한 예상 점수 임계값이 0.5로 설정되었습니다.