

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

• 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건 을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer 🖃





공학석사학위논문

본보기 기반 열린집합 총괄적 분할

Exemplar-Based Open-Set Panoptic Segmentation

2021 년 8월

서울대학교 대학원 전기·정보공학부 황 재 동

본보기 기반 열린집합 총괄적 분할

지도교수 한보형

이 논문을 공학석사 학위논문으로 제출함 2021 년 8 월

> 서울대학교 대학원 전기·정보공학부 황 재 동

황재동의 공학석사 학위논문을 인준함 2021 년 8 월

위 원 장이 경 무부위원장한 보형위 원장병탁

요약

본 논문에서는 총괄적 분할 문제를 열린집합 문제로 확장한 열린집합 총괄적 분할(open-set panoptic segmentation, OPS) 문제를 소개한다. 이 문제는 학습 과정에서 배우는 아는 클래스(known class) 뿐만 아니라 학습 중에는 인지하지 못했던 모르는 클래스(unknown class)들에 대해서도 총괄적 분할을 하는 것을 요구한다. 열린집합 총괄적 분할의 현실적인 어려움에 대해 살펴보고, 총괄적 분할 데이터셋 중하나인 COCO를 이용하여 열린집합 총괄적 분할 문제를 위한 벤치마크를 만든다. 또한, 본보기 기반 이론에서 착안하여, 본보기 기반 열린집합 총괄적 분할 네트워크 (exemplar-based open-set panoptic segmentation network, EOPSN)를 제시한다. 이 방법은 클러스터링을 통해서 식별되고, 유사정답으로 사용되는 본보기들을 기반으로 새로운 클래스를 찾으며, 이미 식별된 본보기들을 이용하여 각 클래스들의 크기를 증가시킨다. EOPSN의 성능을 확인하기 위해 제안된 벤치마크에서 평가를 하여 효과를 입증하였다. 이 연구의 주요 목표는 연구자들에게 열린집합 상황에서의 인식 문제들에 대한 관심을 증대시키는 데에 있다.

주요어: 열린집합, 총괄적 분할, 본보기 기반 학습, 클러스터링, 마이닝

학번: 2019-29143

목 차

ያ	약		ii
1	서	론	1
2	관련	. 연구	5
	2.1	총괄적 분할	5
	2.2	열린집합 학습	5
	2.3	본보기 기반 학습	6
3	열린	집합 총괄적 분할	8
	3.1	레이블과 문제의 정의	8
	3.2	메트릭	9
	3.3	문제점	10
	3.4	다루기 쉬운 태스크 정의	11
4	방	법	12
	4.1	동기와 개요	12
	4.2	베이스라인	13
	4.3	본보기 기반 학습	13
5	실험	결과	17
	5.1	데이터셋과 평가	17
	5.2	구현 및 학습	18
		5.2.1 구형	18

		5.2.2 학습	18
	5.3	정량적 결과	19
	5.4	하이퍼파라미터	20
	5.5	정성적 결과	20
6	분	석	27
7	결	론	28
Ab	strac	t :	35
간	사의	-	36

표 목차

丑 5.1	COCO val 셋 $(K=20\%)$ 에서 다양한 베이스라인들에 대한	
	비교실험	23
丑 5.2	COCO val 셋에서 다양한 아는-모르는 클래스 스플릿에 대	
	한 열린집합 총괄적 분할 결과	24
丑 5.3	$\operatorname{COCO}\mathit{val}$ 셋 $(K=10\%)$ 에서 다양한 하이퍼파라미터에 대	
	한 민감도 분석	25

그림 목차

그림 1.1	예시	2
그림 3.1	이미지에 대한 총괄적 분할에 대한 예측(prediction)과 정답(ground-truth)	9
그림 4.1	제안하는 프레임워크(EOPSN)에 대한 그림	
그림 5.1	EOPSN에서 첫번째 클러스터링 과정에서 식별된 모르는	
	클래스에 대한 본보기들	21
그림 5.2	COCO val 셋에서의 정량적 결과 $(K=20\%)$	22
그림 5.3	$\operatorname{COCO}\mathit{val}$ 셋에서의 정량적 결과($K=10\%$)	26

제1장서 론

연구자들은 컴퓨터 비전 연구를 더욱 깊이있게 하기 위해 대규모의 이미지 기반 벤치마크를 만들고 [5,17,22,30,41] 이미지 인식(image recognition) [14,45], 물체 검출(object detection) [9,21,39,40], 의미론적 분할(semantic segmentation) [3,25,31], 객체 분할(instance segmentation) [2,13] 등과 같은 다양한 과제들에 대해 연구해왔다. 최근 연구자들은 객체(instance)들의 위치와 모양을 찾는 것과 자연스러운 풍경에서 의미있는 레이블(label)들을 예측하는 것에 대해 더 관심을 기울이고 있는데, 이러한 것들은 시각적인 컨텐츠들을 더욱 고차원적으로 이해하는데 중요하기 때문이다.

Kirillov와 동료들 [16]이 최근 객체 분할과 의미론적 분할을 합친 총괄적 분할 (panoptic segmentation) 과제를 만든 이후로 다양한 연구들 [4,8,15,18,20,23,36,46,50] 이 수행되었으며 상당한 성과를 거두고 있다. 하지만, 총괄적 분할의 한가지 문제점은 데이터셋을 만드는 데 엄청난 비용이 든다는 것이다. 픽셀 레벨의 총괄적 분할 정답(ground-truth)을 만드는 일은 상당한 양의 노력을 필요로 하며, 의미론적 분할 보다 더욱 어려워서 큰 스케일의 데이터셋을 만드는 것은 무척 어려운 일이다. 이를 해결하기 위해 두 가지의 방법이 있다. 첫째로는 학습할 때 이러한 정답들을 필요로 하지 않는 약한지도학습(weakly-supervised learning) 하의 총괄적 분할 모델 [19]을 만드는 것이다. 다른 방법으로는 학습된 클래스에 대한 영역분할을 하는 것 뿐만 아니라 학습 때 보지 못한 클래스(unknown class)들에 대해서도 영역분할을 할 수 있는 일반화된 모델을 만드는 것이다. 두번째 방법은 알지 못하는 클래스가 평가 때 등장하는 것을 허용하는 열린집합 인식(open-set recognition) [42]과 밀접한 관련이 있다.

본 논문에서는 총괄적 분할을 열린집합 문제로 확장하였다. 기존의 닫힌집합 총

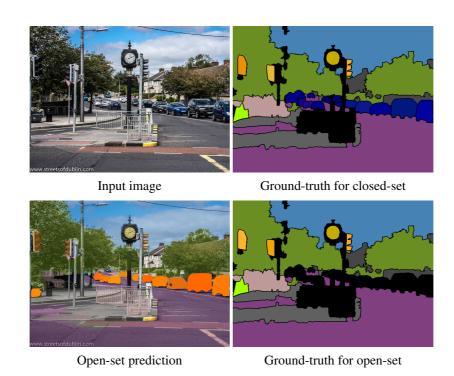


그림 1.1: 주어진 이미지에 대해 닫힌집합과 열린집합 총괄적 분할에 대한 정답들이며 레이블이 없는 부분은 검은색으로 표시하였다. 열린집합 총괄적 분할에서 모델은 학습 데이터에 정답이 없는 모르는 클래스에 대한 객체들을 찾아야 한다. 여기에서는 "자동차"가 모르는 클래스이다.

괄적 분할과 달리 이 새로운 태스크는 학습 때 인지하지도, 학습되지도 않은 모르는 클래스(unknown class)에 대해서도 다루어야 한다. 그러므로 이 태스크의 목표는 아는 클래스(known class)와 모르는 클래스(unknown class) 모두에 대해 총괄적 분할 레이블을 찾는 것이다. 이러한 태스크를 열린집합 총괄적 분할(open-set panoptic segmentation, OPS) 이라고 부른다. 이는 보다 복합적인 시각 이해를 위한 중요한 마일스톤이며 데이터셋 생성, 장면 분석, 이미지 편집 등 다양한 응용분야에 상당한 영향을 끼칠 수 있다. 그림 1.1 는 OPS의 예를 보여준다.

OPS로 나아가기 위해서는 여러가지 어려움이 따르는데, 이를 다루고 몇 가지

합리적인 가정들을 통해 풀 수 있는 문제로 바꾸었다*. 가장 첫번째 어려움은 *물체*의 정의에서 온다. 특정 물체는 많은 물체들의 집합으로 구성이 될 수 있기 때문에 하나의 클래스의 물체는 여러 개의 하위 범주들로 나뉠 수 있다. 예를 들어 "버스"는 창, 사이드 미러, 바퀴 등과 같은 다양한 부품으로 구성된다. 한편, "풀"이나 "물"과 같이 *stuff*에 속하는 클래스들은 지역으로 정의되거나 개별적인 객체를 정의하기 어렵기 때문에 알 수 없는 *stuff* 클래스들을 다루는 것은 어렵다. 따라서 모든 종류의 물체(*thing*)와 의미있는 배경(*stuff*)에 대한 철저한 총괄적 분할은 거의 불가능하다. 두 번째 문제는 알 수 없는 클래스의 이미지를 구별하는 열린집합 이미지 인식과 달리, 열린집합 총괄적 분할은 알 수 없는 클래스 객체가 배경(background)이라고 레이블 되어있어도 이미지에서 알 수 없는 클래스에 대한 객체들을 찾아야 한다는 것이다.

본 논문에서는 열린집합 총괄적 분할 태스크를 위해 본보기 기반 열린집합 총괄적 분할 네트워크(exemplar-based open-set panoptic segmentation network, EOPSN)라는 뛰어난 베이스라인(baseline)을 제안하였다. 유명한 닫힌 집합 총괄적 분할을풀기 위한 모델 중 하나인 Panoptic FPN [15]을 바탕으로 심리학의 본보기 이론(exemplar theory) [27, 32]에서 착안하여 모르는 클래스를 찾는 추가적인 요소를 통합하였다. 우선 모르는 클래스와 이에 대한 본보기(exemplar)들을 물체 특징(object feature)들을 클러스터링(clustering)하여 찾고, 학습 중에 이미 찾은 본보기들과의유사성에 기반하여 더 많은 본보기들을 찾는다. 제안된 모델은 일반적이며 다른 하향식(top-down) 총괄적 분할 방법들에 적용할 수 있다. 여러 실험결과들은 제안된모델, 즉 EOPSN이 Panoptic FPN [15]을 변형한 단순한 베이스라인들보다 우수함을 증명한다. 이 논문의 기여는 아래와 같다.

• 열린집합 총괄적 분할(open-set panoptic segmentation, OPS)을 정의하고, 내 재적인 어려움들에 대한 심층적인 분석을 토대로 합리적인 가정들을 세워 풀

^{*3.3}절에서 상세하게 다룬다.

수 있는 문제로 만들었다.

- COCO [22]를 이용하여 새로운 OPS 벤치마크를 만들고, Panoptic FPN에 기 반한 다양한 베이스라인들의 성능을 보였다.
- 총괄적 분할을 위해 새로운 프레임워크인 EOPSN을 본보기 이론에 기반하여 만들고, 이것이 모르는 클래스들을 잘 찾고, 영역분할을 잘 하는 것을 보였다.

제 2 장 관련 연구

2.1 총괄적 분할

의미론적 분할과 객체 분할 문제를 합친 총괄적 분할은 Kirillov와 동료들 [16] 이 소개한 이후로 많은 관심을 받고 있다. 이 문제에 대한 많은 작업이 존재하며 하향식(top-down)과 상향식(bottom-up) 접근법의 두 그룹으로 분류된다. 하향식 방법들 [8, 15, 18, 20, 23, 36, 50]은 일반적으로 물체 제안(object proposals)을 만들고, 객체 분할한 것을 의미론적 분할 결과와 합친다. 주로 Mask R-CNN [13]이 객체 분할에 쓰이며, 인코더-디코더 구조가 의미론적 분할에 이용된다. 이 위에 AUNet [20]은 마스크 수준의 어탠션(attention)을 이용하여 객체 분할 헤드(instance head)의 지식을 의미론적 분할 헤드(semantic head)로 전달한다. UPSNet [50]는 파라미터가 필요없는(parameter-free) 총괄적 헤드(panoptic head)를 두어 상충되는 물체(thing)와 의미있는 배경(stuff) 예측의 해결한다. 한편, 일반적으로 제안이 필요 없는(proposal-free) 상향식 방법들은 의미론적 분할을 먼저 하고, 객체들로 쪼갠다 [4,46,51]. 상향식 방법들은 객체-객체 혹은 객체-배경의 중첩 문제로부터 자유롭지만 일반적으로 하향식 방법들에 비해 성능은 떨어진다. 본 논문에서는 객체(thing)와 의미있는 배경(stuff)을 분리하여 처리하는 하향식 방법을 사용하였다.

2.2 열린집합 학습

열린집합 인식(open-set recognition)은 컴퓨터 비전 사회에서 최근 몇 년 동안 주목을 받고 있다. 열린집합 이미지 분류의 목표는 모르는 클래스들에 대한 이미지를 인식하고, 학습에 사용한 아는 클래스(known class)에 대한 이미지들을 분류하는 것이다 [42]. 열린집합 분류 알고리즘 [43]에서 모르는 클래스(unknown class)들은 학습 때 부정적인 클래스(negative class)로 사용되었는지(본 모르는 클래스,

seen unknown class)와 보지 않았는지(보지 않은 모르는 클래스, unseen unknown class)로 나뉜다. 한편, 열린집합 객체 검출(open-set object detection) [7]처럼 학습데 이터에서 레이블이 안되어있는 물체들이 등장하였는지 여부를 할 수 없기 때문에 열린집합 총괄적 분할 문제에서는 이러한 구분을 하기 어렵다.

대부분의 열린집합 이미지 분류 방법들 [1,24,29,34]은 테스트 때 아는 클래스에 대한 확률이 특정 문턱값(threshold)보다 낮은 경우 모르는 클래스라고 예측한다. OpenMax [1]는 (C+1) 가지 분류기를 배경과 C- 클래스 분류에 사용하며, 분류기를 Weibull 분포를 이용하여 학습한다. 이미지 생성(image generation) [34]이나조건법적 이미지(counterfactual image)들 [29]이 아는 클래스들을 위한 강건한 결정경계를 예측하고, 효과적으로 모르는 클래스를 찾는데 이용되기도 한다. Liu와 동료들 [24]은 아는 클래스의 분포가 긴 꼬리 분포(long-tailed distribution)인 상황에서의열린집합 이미지 분류에 메모리와 클러스터링 기반의 모델을 제시하였다.

최근에는 열린집합 시나리오가 물체 검출 [7,28]이나 의미론적 객체분할(semantic instance segmentation) [35] 등의 분야로 확장되었다. Miller와 동료들 [28]은 드롭아웃 추출(dropout sampling)을 이용하여 열린집합 물체 검출을 처음 제시하였으며, Dhamija와 동료들 [7]에 의해 더 발전되었다. Pham과 동료들 [35]은 물체의 경계와 마스크들을 이용하여 열린집합 객체 영역 분할을 위한 베이지안 최적화 프레임워크를 제안하였다. 본 논문에서는 열린집합 총체적 분할 문제를 닫힌 집합 문제를 기반으로 만들어 냈다.

2.3 본보기 기반 학습

클러스터의 중심들을 각 클러스터의 대표들로 여기는 원형 기반 방법(prototype-based method)들 [44, 49]과는 달리 본보기 기반 방법(exemplar-based method)들은 메모리에 저장된 샘플들을 곧바로 사용한다. 기계학습에서 대표적인 본보기 기반 방법으로는 새로운 데이터의 레이블을 이웃한 학습 데이터의 레이블을 이용해서 붙이는 k-최근접 이웃 알고리즘(k-nearest neighbor algorithm) [6]이 있다. Wu와 동

료들 [48]은 각각의 예시들을 개별적인 클래스로 취급해서 어노테이션 없이 표상학습(representation learning)을 진행하였다. 점진적 학습(incremental learning) 방법들 [12, 37, 38]은 종종 본보기 기반 학습을 파괴적 망각(catastrophic forgetting)을 막는데 사용한다. 이 방법들은 전체 데이터 셋이나 기존 태스크의 클래스들을 효과적으로 대표하는 소수의 본보기들을 저장한다. MemAE [10]는 본보기 기반 학습을이상 탐지(anomaly detection)에 사용하였다.

기존의 본보기 기반 학습 방법들은 본보기들에 대한 정확한 정답 레이블들이 존재하거나 그 정보들이 별로 중요하지 않다는 가정을 한다. 하지만 우리의 문제에 서는 다양한 모르는 클래스들에 대한 본보기들을 자동으로 식별하고 레이블 없이 그들의 표상(representation)들을 배워야 하므로 훨씬 어려운 문제이다.

제 3 장 열린집합 총괄적 분할

이 장에서는 열린집합 총괄적 분할(open-set panoptic segmentation, OPS)과 평가 기준에 대해 정의하였다. 또한, 이 문제가 갖고 있는 본질적인 어려움들을 논하고, 합리적인 가정들을 통해 어떻게 풀 수 있는 문제로 바꾸었는지 다루었다.

3.1 레이블과 문제의 정의

열린집합 총괄적 분할은 닫힌집합 총괄적 분할과 비슷하다. 열린집합 총괄적 분할에서는 학습과 테스트 때 모두 존재하는 아는 클래스(known class)뿐만 아니라 학습에 사용하지 않았으나 테스트할 때 등장하는 모르는 클래스(unknown class)가 존재한다. 클래스를 분류하는 또 다른 방법으로는 물체가 물리적으로 어떻게 형성되었는지이다. thing 클래스, C^{Th} 는 구체적인 형상을 지니고 있는데에 반해, stuff 클래스, C^{St} 는 비정형의 배경(예: "하늘"이나 "모래") 또는, 구조가 불분명한 물체(예: "나무", "풀")들로 이루어져 있다. 또한, 모호하거나 범주에 없어서 학습 데이터에 레이블이 안 되어 있는 영역들을 의미하는 레이블인 void 클래스가 존재한다.

미리 정의된 C개의 아는 클래스들의 집합이 $C := \{0,...,C-1\}$ 로 주어지고, 모르는 클래스의 집합이 U라고 주어졌을 때 이미지의 i 번째 픽셀은 튜플 $(l_i,z_i) \in$ $(C \cup U) \times \mathbb{N}$ 로 레이블링 되며, l_i 와 z_i 는 각각 의미 클래스 레이블과 객체 식별자이다. 본 문제에서 U는 모르는 클래스 전체를 대표하는 한개의 원소로 이루어져 있다. 같은 객체 식별자 (z_i) 를 갖고 있는 픽셀들의 그룹은 클래스 $l_i \in (C \cup U)$ 를 나타내는 하나의 영역이 된다. 요약하자면, 열린집합 총괄적 분할은 주어진 이미지로부터모든 아는 클래스들과 모르는 클래스들의 영역을 분할하는 것을 목표로 한다.

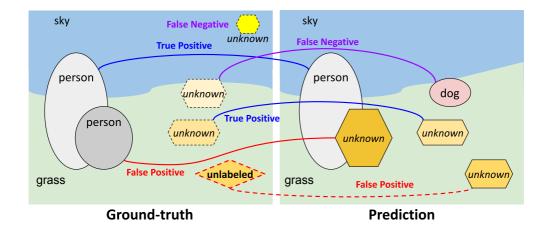


그림 3.1: 이미지에 대한 총괄적 분할에 대한 예측(prediction)과 정답(ground-truth)을 간략히 표시한 그림. Ground-truth에서 검은 점선으로 표시된 물체들은 모르는 클래스들을 의미하며, 빨간 점선으로 표시된 물체는 레이블이 없는 물체이다. 두 물체 사이의 연결은 매칭이 된 것이다. 어떤 모델이 우하단의 모르는 객체를 정확히 찾았다고 하더라도 이는 ground-truth에 레이블이 없는 객체여서 거짓 양성으로 여겨지게 된다.

3.2 메트릭

아는 클래스들과 모르는 클래스들에 대해 총괄적 분할에서 일반적으로 사용하는 메트릭(metric) [16]인 총괄적 품질(panoptic quality, PQ), 분할품질(segmentation quality, SQ), 인식품질(recognition quality, RQ)을 사용하였다. 이 메트릭들은 아래와 같이 정의된다.

$$PQ = \underbrace{\frac{\sum_{(p,g) \in TP} IoU(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \cdot \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}.$$
(3.1)

이 때, IoU는 두 영역의 교집합 대 합집합의 비(Intersection-over-Union)이고, 그림 3.1에서처럼 TP, FP, and FN는 각각 참 양성(true positive), 거짓 양성(false positive), 그리고 거짓 음성(false negative)에 대한 집합을 의미한다. 만약 예측된 분할영역이 같은 레이블의 참 영역과 일정 정도 이상 중첩(즉, IoU > 0.5)되어 있으면, 이 영역은 참 영역 집합, TP으로 분류된다.

3.3 문제점

모르는 객체들 찾기 단순히 모르는 클래스에 대한 이미지들을 인식하면 되는 열린 집합 이미지 인식(open-set recognition)는 달리 열린집합 총괄적 분할 모델은 각각의 이미지에서 모르는 클래스에 대한 물체들을 찾아야 한다. 이러한 물체들에 대해 사전지식이 없고, 모르는 물체들은 학습 중에는 배경(background)으로 레이블이 되었을 것이기 때문에 이는 더욱 어려운 문제가 된다.

정답 레이블링 총괄적 분할은 의미가 있는 배경 영역들과 각각의 객체들에 대한 픽셀 수준의 어노테이션(annotation)이 필요하다. 어노테이터(annotator)들은 모든 객체와 stuff 들을 철저하게 표시해야한다. 하지만, 어떤 물체는 여러 작은 물체들로 구성되어 포괄적인 어노테이션을 만드는 것은 매우 어렵다. 예를 들어 "자동차"는 창문, 사이드미러, 조명, 타이어 등과 같은 여러 객체들로 구성된다. 또한, 일부 stuff 레이블들(예: "나무", "자갈")은 여러 객체들로 나눌 수 있다. 구체적인 지침이 없으면 어노테이션들은 일관적이지 않고 신뢰할 수 없게 되며, 이는 열린집합 총괄적 분할 모델을 학습하고 평가하는 것을 저해한다.

평가 총괄적 분할을 측정하기 위한 현재의 메트릭(metric)들은 이미지들이 완전하게 어노테이션 되어있다고 가정한다. 하지만, 클래스 레이블들의 우주 집합(universe set)은 실제로는 무한하여, 총괄적 분할의 실제 성능과 평가가 불일치하게 만든다. 즉, 그림 3.1과 같이 열린집합 총괄적 분할 알고리즘은 모르는 클래스에 대한 객체들을 매우 정확하게 식별하였는데, 이에 대한 레이블이 정답에 없을 경우, 거짓 양성이 증가하여 RQ와 PQ를 낮추게 된다. 한편, SQ는 참 양성들만 고려하여 이에 대한 영향을 받지 않으나, 전반적인 총괄적 분할의 성능을 나타내지 못한다는 한계를 지닌다.

3.4 다루기 쉬운 태스크 정의

3.3절에서 살펴본 바와 같이 OPS의 개념적이고 실제적인 명백한 문제들로 인해 아래와 같이 다루기 쉬운 버전의 문제로 정의하였다.

가정 학습 데이터에서 등장하는 모르는 클래스에 대한 세 가지 가정을 하였다. 첫째, 모든 모르는 클래스들은 thing 카테고리에 속한다. 이는 stuff 카테고리에 모르는 클래스가 있을 경우 종종 정의가 불분명하기 때문이다. 두 번째 가정은 아는 클래스의 일부분은 모르는 클래스가 될 수 없다는 것이다. 예를 들어 "자동차"와 "사람"이 아는 클래스라고 했을 때, 바퀴나 머리는 아는 클래스의 일부분이기 때문에 모르는 클래스에 대한 객체가 될 수 없다. 하지만 만약에 바퀴가 자동차의 일부분이 아니라홀로 등장하였을 경우에는 모르는 클래스가 될 수 있다. 마지막 가정은 모르는 클래스의 객체들은 학습 때 void 영역에만 등장한다는 것이다. 위의 가정들의 목적은 아는 클래스의 영역과 모르는 클래스의 영역 사이의 혼동을 막기 위함이다.

데이터셋 열린집합 총괄적 분할을 위한 완벽한 어노테이션은 불가능하다. 그래서 OPS의 성능을 측정하는 합리적인 방법은 여러 도메인 의미론적 분할(multi-domain semantic segmentation)처럼 하나의 데이터셋에서 학습하고, 모르는 클래스가 있는 다른 데이터셋에서 테스트하는 것이다. 하지만, 데이터셋 간의 도메인 격차(domain gap)이 있고, 물체에 대한 정의는 데이터셋마다 다르다. OPS를 위해 존재하는 한 가지 데이터셋만을 이용했다. 총괄적 분할을 위해 COCO [22], Cityscape [5], Mapillary [30], ADE20k [52] 등과 같은 여러 벤치마크들이 있다. 하지만 COCO를 제외한 다른 데이터셋들은 자동차 운전이나 실내와 같이 특정 환경에 전문화되어 있어서 일상 생활에 대한 사진들을 많이 담고 있는 COCO를 사용하였다. thing 클래스에서 다른 개수의 모르는 클래스들을 추출하여 여러 스플릿(split)을 만들었다. 그러나 레이블 정보가 학습 및 및 평가 전반에 걸쳐 완전히 누락 되었기 때문에 본 논문에서 제안하는 벤치마크는 앞서 설명한 것처럼 모르는 클래스를 제대로 평가할 수 없다.

제4장방 법

이 장에서는 열린 집합 총체적 분할을 위한 베이스라인 모델들과 제안하는 방법을 소개한다.

4.1 동기와 개요

OPS에서 핵심이 되는 요소는 어떻게 모르는 클래스들에 대한 객체들을 찾을 것인가이다. 한 가지 방법으로는 물체 점수(objectness score)를 예측하는 RPN [40] 과 같은 클래스에 무관한(class-agnostic) 모델을 사용하는 것이며, 이를 본 논문의 베이스라인으로 삼았다. 하지만, 물체 점수가 아는 클래스들에 최적화되어있다는 한계를 지닌다. 그러므로 모델의 성능은 아는 클래스와 모르는 클래스가 얼마나 의미적으로 비슷한지에 크게 의존한다. 즉, 어떤 아는 클래스와도 의미적으로 비슷하지 않은 모르는 클래스는 찾기 힘들다.

이 문제를 해결하기 위해 학습 데이터로부터 모르는 클래스들을 보다 효율적으로 찾을 수 있는 본보기 기반 학습 프레임워크를 베이스라인 위에 적용하였다. 여기에서 본보기란 모르는 클래스로 식별되는 바운딩 박스(bounding box)를 의미한다. 이 모델은 우선 k-중심 클러스터링(k-means clustering) [26]을 이용하여 모르는 클래스들과 그에 해당하는 본보기들을 void 영역에 있는 식별된 바운딩 박스들에서 찾는다. 높은 물체 점수를 갖고 있는 빽빽한 클러스터는 모르는 클래스로 여겨지며, 클러스터 내의 각 이미지 패치는 본보기로 여겨진다. 이 본보기들은 후속 미니배치 (mini-batch)에서 이미지 내의 영역제안들과 유사도를 계산하여 새로운 본보기들을 찾는데 이용된다. 또한, 본보기들은 물체 인식에 대한 유사 정답(pseudo-ground-truth)으로도 이용된다. 클러스터링과 마이닝 단계들에서는 백본 네트워크가 아는 클래스들과 모르는 클래스들 모두에 최적화되는 동안, 새로운 모르는 클래스들을

찾고 식별된 모르는 레이블에 대한 본보기들을 모은다. 그림 4.1은 제안된 방법의 단계들을 나타낸다.

4.2 베이스라인

앞서 소개한 문제의 정의에 의해 오직 thing 클래스들만이 모르는 클래스가 될수 있기 때문에 RPN을 이용한 하향식 총체적 분할 방법을 적용하는 것을 베이스라인 (baseline)으로 삼는 것은 합리적이다 객체 분할 헤드(instance head)와 의미론적 분할 헤드(semantic head)가 특징 추출기(feature extractor)를 공유하며 그 위에 존재하는 Panoptic FPN [15]을 이용하였다. 객체 분할 헤드는 Mask R-CNN [13]과 정확히 일치하며, 의미론적 분할 헤드는 FPN 기반의 디코더(FPN-based decoder)이다.

모르는 클래스들을 다루기 위하여 객체 분할 헤드의 바운딩박스 리그레서 (bounding box regressor)와 마스크 예측기(mask predictor)를 클래스와 무관(classagnostic)하게 변경하였다. 이 베이스라인 모델은 아는 클래스에 대한 바운딩 박스들을 먼저 예측하고, RPN에서 만들어진 후보들의 물체 점수와 문턱값 0.5를 이용하여 모르는 클래스들에 대해서도 찾는다. void 영역에서 샘플된 바운딩 박스들을 식별하기 위해 객체 분할 헤드의 분류기에서 void 클래스를 학습하였다. void 클래스에 대한 지도(supervision)를 위해서 영역의 절반 이상이 void 영역에 있는 바운딩 박스들을 이용하였다. void 영역이 물체를 나타낸다는 보장을 할 수 없으므로, RPN의 학습에는 사용하지 않았다.

4.3 본보기 기반 학습

본보기 기반 학습(exemplar-based learning)의 주요 목표는 모르는 클래스들에 대한 일관된 특징(feature)들을 갖고 있는 바운딩 박스들을 식별하는 것이다. 본 논문에서 제안하는 본보기 기반 학습은 두 단계로 이루어져 있다: 새로운 클래스와 본보기들을 찾기 위해 학습 이미지들의 부분집합에서 얻어지는 모르는 클래스들에 대한 영역 제안 후보들을 클러스터링, 이미 찾아 놓은 본보기들과 유사도 비교를

통해서 새로운 본보기들을 마이닝. 이 클러스터링과 마이닝 단계들은 학습 전 과정에서 번갈아 나타난다.

마이닝과 클러스터링 단계들을 위해 모르는 클래스들을 위한 void 영역의 모르는 바운딩 박스들로부터 특징을 추출하였다. 이를 위해 먼저 중복된 검출들을 줄이기 위해 Non-Maximum Suppression (NMS)를 IoU (Intersection over Union) 임계값 1×10^{-7} 으로 하여 적용하고, 제안들을 RPN [40]에서 얻어진 물체점수들을 기반으로 가중추출하였다. 그 후에 백본 네트워크를 이용하여 영역제안들로부터 얻어지는 1024 차원의 특징들을 얻고, [13].와 같이 RoI-Align, GAP 그리고 두개의 완전 연결계층(fully-connected layer)를 이용하여 특징을 계산하였다.

클러스터링(Clustering) 모르는 클래스들을 찾기 위해 후보 바운딩 박스들로부터 얻어진 특징들에 코사인 거리에 기반한 k-중심 클러스터링을 적용하였다. 클러스터링은 매 200 스탭마다 직전 200 스탭동안 샘플된 제안들을 위한 특징들에 적용되었다. 클러스터를 과하게 만들고(over-clustering) 클러스터들의 부분집합만을 모르는 클래스와 매우 잘 일치하는 클러스터들만 사용하였다. 모르는 클래스들을 나타내는 객체들은 긴밀하게 모이는 경향이 있는 반면, 물체를 나타내지 않는 영역 제안들은 서로 연결관계가 약하고, 낮은 물체점수를 지니고 있을 것이기 때문에 모르는 클래스들를 나타내는 클러스터들은 진짜 높은 평균 물체 점수를 지니고 있으면서 중심 (centroid)과 모든 원소들간의 평균 코사인 거리가 작아야 한다. 좋은 품질의 클러스터들에서 본보기들을 식별하고 이어지는 마이닝 단계를 위해서 저장한다. 학습하는 동안 여러 식별된 모르는 클래스들을 위한 본보기들이 각각 존재하지만, 평가할 때는 이를 모두 하나의 클래스로 취급하였다.

마이닝(Mining) 과거에 검출된 모르는 클래스들에 대한 추가적인 본보기들을 이후의 미니배치(mini-batch)의 이미지로부터 찾는다. 새로운 미니배치의 이미지들의물체 제안들의 특징들과 저장된 본보기의 특징들과 비교하여 쉽게 얻어진다. 특징

추출 네트워크가 계속 업데이트가 되기 때문에 저장된 본보기의 특징들을 다시 계산할 필요가 있다. 만약 RPN에서 생성되는 제안과 존재하는 어떤 본보기들과 코사인 유사도가 문턱값보다 높으면 이 제안을 새로운 본보기로 받아들였다.

손실함수(Loss) Panoptic FPN [15]의 회귀 손실함수(regression loss)를 포함한 분류 손실함수(classification loss), 마스크 손실함수(mask loss), 의미론적 분할 손실함수(semantic segmentation loss)와 거의 유사한 손실함수들을 사용하였다. 이 손실함수들과 유일하게 다른 점은 객체 분할 헤드에 있는 분류 손실함수이다. 다음과같이 아는 클래스, 배경 클래스(background class, bg), 모르는 클래스들에 대해서는 크로스 엔트로피 손실함수(cross-entropy loss)를 사용하였다.

$$\mathcal{L}_{ce} = \sum_{c \in (\mathcal{C}^{Th} \cup \{bg\} \cup \mathcal{U})} -y_c \log p_c. \tag{4.1}$$

여기에서 y_c 는 (유사)정답 레이블이고 p_c 는 클래스 c에 대한 소프트맥스(softmax) 스코어이다. 추가로, void 지역에 있는 물체 제안들이 아는 클래스들로 잘못 분류되지 않게 하기 위해 다음과 같이 부정적인 지도를 하였다.

$$\mathcal{L}_{\text{void}} = \sum_{c \in C^{\text{Th}}} -\log(1 - p_c). \tag{4.2}$$

즉, 전체 분류 손실함수는

$$\mathcal{L}_{cls} = \mathcal{L}_{ce} + \mathbb{I}_{void} \mathcal{L}_{void}$$
 (4.3)

이며, 여기에서 \mathbb{I}_{void} 는 void 영역에 위치한 박스들을 나타내는 지시함수이다.

추론(Inference) 본 논문에서 제안하는 모델은 객체 분할과 의미론적 분할 결과를 따로 예측하고, 이를 하나로 합쳐서 총체적 분할 결과를 만들어 낸다. 일반적인 총체적 분할 모델 [15]의 추론 메커니즘과 같은 방식으로 총체적 분할지도를 만든 뒤, 모르는 객체들에 대해서 추가적으로 예측한다.

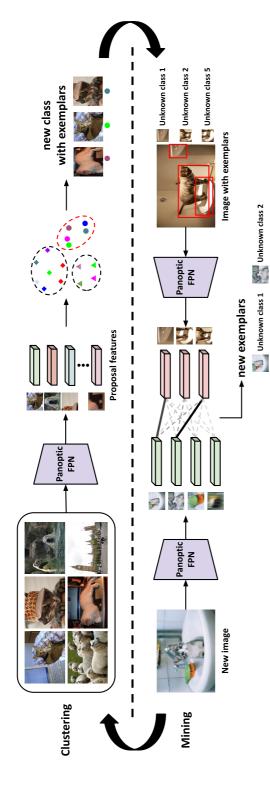


그림 4.1: 제안하는 프레임워크(EOPSN)에 대한 그림. 모델은 우선 모르는 클래스들과 그에 상응하는 본보기들을 이미지들의 작은 부분집합으로부터 void 지역에서 검출된 바운덩 박스(bounding box)들에 대해 k-중심 클러스터링 [26]을 수행하여 찾는다(클러스터 링 단계). 클러스터의 각 원소들은 프로포절 특징(proposal feature)이며 색의 다름은 다른 이미지에서 유래하였음을 의미한다. 빨간색 클러스터는 하나의 모르는 클래스를 나타내며 클러스터 내의 각각의 이미지 패치들은 해당 클래스의 본보기가 된다. 식별된 모르는 클래스들에 대한 본보기들은 마이닝(mining) 단계에서 물체 영역제안들과의 비교를 통해 새로운 본보기들을 찾는다. 클러스터링과 마이닝 단계들은 모르는 클래스를 찾고, 이에 대한 본보기들을 늘리기 위해 수행되는데, 연산을 줄이기 위해 클러스터링은 드물게 시행된다.

제 5 장 실험 결과

본 장에서는 실험 세팅과 실험 결과를 다루었으며, 제안하는 프레임워크를 다각 적으로 분석하였다.

5.1 데이터셋과 평가

모든 실험은 COCO [22]에서 진행하였다. 80개의 thing 클래스와 53개의 stuff 클래스가 있는 2017 총체적 분할 학습 및 검증 스플릿을 사용하였으며, 이미지는 각각 대략 11만 8천장과 5천장이 있다. 열린 집합 상황으로 만들기 위해 학습데이터에서 아는 thing 클래스의 일부에 대한 어노테이션을 지워 이 클래스들을 모르는 클래스로 취급하였다. 모르는 클래스의 비중(5%, 10%, 20%)에 따라서 세 종류의 다른 스플릿을 만들었다*. 스플릿들에서 지워진 클래스들은 아래와 같으며, 누적되어지워졌다.

- 5%: "자동차", "소", "피자", "변기"
- 10%: "보트", "넥타이", "얼룩말", "정지신호"
- 20%: "식탁", "바나나", "자전거", "케이크", "싱크", "고양이", "키보드", "곰"

성능 평가를 위해 일반적인 총체적 분할 메트릭인 PQ, SQ, RQ를 사용하여 아는 클래스와 모르는 클래스들에 대해 각각의 성능을 제시하였다.

^{*}클래스 레이블들을 데이터셋에 등장하는 빈도 순으로 정렬하고 일정 주기로 클래스들을 추출하여 모르는 클래스에 사용할 클래스들을 찾았다.

5.2 구현 및 학습

5.2.1 구현

Panoptic FPN [15]을 기반으로 구현하였으며, ResNet-50-FPN [14,21]을 백본으로 삼았다. Pytorch [33]와 Detectron2 [47]을 이용하였다.

5.2.2 학습

베이스라인 모델은 ImageNet에 미리 학습된 ResNet-50-FPN [14,21]을 이용하였다. 첫 학습률(learning rate)은 0.04이며, 선형 웜업(linear warm-up) [11]을 사용하였고, 30,000와 40,000 스탭 때 학습률에 0.1씩 곱하였다. 전체 학습은 1×라고 불리는 45,000 스탭까지 진행하였다. weight decay는 0.0001이며 모멘텀(momentum)은 0.9였다. 의미론적 분할 헤드는 void 레이블을 하나의 새로운 클래스로 하여 학습하였다. EOPSN은 (4.2)을 이용한 베이스라인 모델(Void-suppression)을 파인튜닝 (fine-tuning)하여 학습시켰으며 이 때 학습률은 0.0004로 30,000 스탭동안 유지하였다. 다른 모든 하이퍼파라미터(hyper-parameters)는 Detectron2 [47]의 기본 값을 사용하였다. 모든 모델은 8개의 Titan V100 GPU를 동기화되는 SGD(synchronized SGD)로 하여 학습하였으며, 각 GPU에는 4장의 이미지가 미니배치를 이루었다.

본보기 기반 학습에서 노이즈를 줄이기 위해 중간과 큰 사이즈의 바운딩 박스들,즉,영역의 넓이가 32^2 보다 큰 박스들만 사용하였다. 각 미니배치에서 최대 20개의물체 영역제안을 추출하였고,매 200 스텝마다 128개의 클러스터들을 만들었다.우수한 품질의 클러스터들을 찾기 위해 중심과 원소들 사이의 코사인 유사도(cosine similarity)의 평균을 기준으로 오직 상위 10%의 클러스터들과 그 원소들을 사용하였다.우수한 품질의 클러스터들을 찾기 위한 물체 점수 문턱값은 0.9에서 시작하여찾은 모르는 클래스의 개수에 따라 0.99까지 선형으로 증가하였다. 클러스터 내에서붙어있는 원소들과 마이닝 단계에서 새로운 본보기를 찾기 위한 코사인 거리(cosine distance) 문턱값들은 각각 0.15와 0.025이며, 0.01까지 서서히 감소하였다.

5.3 정량적 결과

표 5.1는 Panoptic FPN이 void 영역을 사용할 수 있도록 바꾼 다양한 베이스라인 모델이 COCO val 셋에서 모르는 클래스가 20%일 때의 열린집합 총체적 분할 결과를 나타낸다. void 영역에 존재하는 박스들을 사용하는 방법으로는 다음과 같이 4 가지가 있다: 배경 (background)으로 취급하기(Void-background), 학습에서 제외시키기(Void-ignorance), 아는 클래스로 잘못 분류되는 것을 막기 위해 (4.2)을 이용하여 아는 클래스 쪽에 부정적인 지도를 하기 (Void-suppression), 일반적인 크로스 엔트로피 손실함수를 이용하여 새로운 클래스로 취급하여 학습하기(Void-train). 모든 케이스들은 아는 클래스들에 대해서는 비슷한 성능을 보였으나, 모르는 클래스에 대해서는 상의한 결과를 보였다. Void-ignorance 모델이 가장 안좋은 성능을 보였는데, 이는 모르는 클래스에 대한 객체들이 아는 클래스로 잘못 분류되었기 때문이다. void 영역에 위치한 박스들이 RPN에서 배경으로 학습되지 않아 낮은 물체화 점수를 가지지 않기 때문에 Void-suppression와 Void-train 모델들은 다른 두 경우보다 뛰어난 비슷한 성능을 보였다.

표 5.2는 COCO val 셋에서 아는-모르는 클래스에 대한 스플릿을 다르게 하여 실험한 결과이다. 지도 모델(supervised model)은 모르는 클래스 없이 모든 클래스에 대해 학습한 Panoptic FPN이다. 베이스라인(baseline)은 Table 5.1에서 소개한 Voidtrain 옵션을 이용한 Panoptic FPN이다. EOPSN은 아는 클래스에 대해서는 비슷한 성능을 유지하면서도 모든 측면에서 모르는 클래스에 대해 우수한 성능을 보였다. 전반적으로 모르는 클래스에 대한 PQ와 RQ는 아는 클래스에서보다 훨씬 적었던 반면, SQ는 아는 클래스와 모르는 클래스 모두에 대해 비슷하였다. 이는 SQ가 참양성에 대해서만 계산을 하기 때문이다.

5.4 하이퍼파라미터

클러스터의 개수, 클러스터링을 하는 간격, 사용하는 바운딩 박스의 사이즈와 같은 다양한 하이퍼파라미터에 대해서 실험하였으며 K는 10%로 고정하여 진행하였다. 하나의 하이퍼파라미터를 실험하는 동안 나머지 파라미터는 5.2.2 항에서 언급한대로 사용하였다.

클러스터의 개수 표 5.3a는 클러스터의 개수가 성능에 미치는 영향을 보여준다. 모든 경우에 베이스라인(PQ: 8.5, RQ: 73.2, SQ: 11.6)보다 우수한 성능을 보였다. 또한, 작은 개수의 클러스터를 사용할 경우 각 클러스터마다 노이즈의 개수가 늘어 나고, 많은 개수를 사용할 경우 클래스 내 다양성(intra-class variability)이 감소하게 되어 성능이 하락하였다.

클러스터링 간격 표 5.3b는 클러스터링 간격이 짧을수록 RQ와 PQ가 감소함을 보였다. 이는 클러스터링을 할 때 사용되는 물체 영역제안과 이미지의 수가 줄어들어서 클러스터링 단계가 노이즈에 보다 취약하고, 모르는 클래스와 본보기들을 정확하게 찾기 힘들기 때문이다.

물체 영역제안 크기 표 5.3c에 의하면 크거나 중간 사이즈의 제안들을 사용할 때 COCO val 셋에서 K가 10%인 상황에서 가장 좋은 결과를 보였다. 반면, 작은 크기의 물체 영역제안은 클러스터링을 할 때 노이즈를 증가시켜서 모르는 클래스를 정확히 찾는 것을 방해하여 성능이 큰 폭으로 하락하였다.

5.5 정성적 결과

그림 5.1는 K=20%일 때 EOPSN에서 첫번째 클러스터링을 하고 나서 찾은 본보기들이다. 대부분의 본보기들은 "자동차"를 나타내고 있으며, 그 외에 "케이크",



그림 5.1: EOPSN에서 첫번째 클러스터링 과정에서 식별된 모르는 클래스에 대한 본보기들.

"소", "자전거" 등의 모르는 클래스를 나타내거나 아는 클래스인 "기린"에 대한 객체를 나타낸다.

그림 5.2은 COCO val 셋에서 K=20%일 때, 베이스라인 모델(세번째 줄)과 EOPSN(네번째 줄)에 대한 결과이다. 그림에 모르는 클래스들은 "정지 신호", "자동차", "키보드", "싱크", "변기"가 있다. 두번째 줄은 모르는 클래스가 오랜지색으로 표시된 정답이다. EOPSN은 베이스라인이 놓친 모르는 클래스들을 많이 찾았으며, 키보드의 자판이나 욕조와 같이 COCO에 없는 클래스들도 찾았는데 이들은 모두 거짓 양성으로 계산되어 PQ와 RQ를 낮춘다. 그림 5.3는 COCO val 셋에서 K=10%일 때의 결과이다.



그림 5.2: COCO val 셋에서의 정량적 결과(K=20%). 첫 번째 열은 인풋 이미지들이고, 이어지는 열들은 각각 정답, 베이스라인 $(Void\text{-}train})$ 의 결과, EOPSN의 결과를 나타낸다. 모르는 클래스에 대한 객체들은 주황색으로 표시하였다.

표 5.1: COCO val 셋(K = 20%)에서 다양한 베이스라인들에 대한 비교실험. 베이스라인들은 void 영역을 학습 과정에서 사용하는 방법에 따라, Void-background (배경으로 사용), Void-ignorance (학습에서 제외), Void-suppression (식 (4.2)을 이용하여 학습), Void-train (새로운 클래스로 학습)으로 나뉜다.

					Known	u				Ö	Unknown	"u
)	PQ	SQ	RQ	PQ^{Th}	$\mathrm{SQ}^{\mathrm{Th}}$	RQ^{T}	h PQ St S	SQ St F	$\mathrm{RQ}^{\mathrm{St}}$	PQ	SQ	RQ
Void-background	37.7	76.3	46.6	44.8	79.3	54.1	29.2	72.8	37.5 4.0 71.1	4.0	71.1	5.7
Void-ignorance	37.2	76.3	45.9	43.9	79.0	53.1	29.1	73.0	37.3 3.7	3.7	71.8	5.2
Void-suppression	37.5	75.9	46.1	45.1	9.08	54.5	28.2	70.2	36.1	36.1 7.2	75.3	9.6
Void-train	36.9	76.4	45.5	44.0	80.3	53.3	28.2	71.7	36.0	7.8	7.8 73.4	10.7

표 5.2: COCO val 셋에서 다양한 아는-모르는 클래스 스플릿에 대한 열린집합 총괄적 분할 결과. K는 전체 클래스에서 모르는 클래스의 비율을 의미한다. 비교대상들보다 높은 값들은 굵은 글씨로 표기하였다.

(%) A	Model					Known	u				Uı	Unknown	u
(0/) 🕶	IMORCI	PQ	SQ	RQ	$\mathrm{PQ}^{\mathrm{Th}}$	$\mathrm{SQ}^{\mathrm{Th}}$	SQ RQ $\mid PQ^{Th} \mid SQ^{Th} \mid RQ^{Th} \mid PQ^{St} \mid SQ^{St} \mid RQ^{St} \mid PQ \mid SQ \mid RQ$	PQ St	SQ St	RQ St	PQ	SQ	RQ
	Supervised	39.4	7.77	48.4	39.4 77.7 48.4 45.8 80.7	80.7	55.4	55.4 29.7 73.1	1	38.0	ı	ı	ı
v	Baseline (<i>Void-train</i>) 37.7 76.7 46.4 44.2	37.7	76.7	46.4	44.2	80.4	53.5 28.3	28.3	71.3	71.3 36.2 10.0 73.8 13.5	10.0	73.8	13.5
J	EOPSN	38.0	76.9	46.8	38.0 76.9 46.8 44.8	80.5	54.2	28.3	71.9	36.2 23.1 74.7 30.9	23.1	74.7	30.9
10	Baseline (Void-train) 36.9 75.4 45.5 43.2	36.9	75.4	45.5	43.2	79.0	52.4 28.3	28.3	70.4	70.4 36.2 8.5 73.2 11.6	8.5	73.2	11.6
10	EOPSN	37.7	37.7 76.8 46.3	46.3	4.5	80.6	53.8	28.4	71.8	71.8 36.2 17.9	17.9	8.9/	23.3
6	Baseline (Void-train) 36.9 76.4 45.5 44.0	36.9	76.4	45.5	44.0	80.3	53.3	28.2	71.7	71.7 36.0 7.8 73.4 10.7	7.8	73.4	10.7
07	EOPSN	37.4	37.4 76.2 46.2	46.2	45.0	80.3	54.5	28.2	71.2	36.2 11.3	11.3	73.8	15.3

표 5.3: COCO val 셋(K=10%)에서 다양한 하이펴파라미터에 대한 민감도 분석.

(c) 물체 영역 제안의 크기	Honora Targe Known Unknown Earge RQ	_
(b) 클러스터링 간격	플러스터링 간격 Rnown Unknown 100 37.6 77.2 46.3 8.2 77.5 10.6 200 (ours) 37.7 76.8 46.3 17.9 76.8 23.3 400 37.7 77.5 46.3 14.6 76.4 19.1	
(a) 클러스터 개수	클러스터 개수 <mark>아는 클레스 R로는 클레스 RQ </mark>	



그림 5.3: COCO val 셋에서의 정량적 결과(K=10%). 각 행은 이미지, 정답, EOPSN의 결과를 의미한다. 모르는 클래스에 대한 객체들은 주황색으로 표시하였다.

제6장분 석

3.3 절에서 논의한 바와 같이, 여러가지 중요한 문제점들이 남아있다. 우선, EOPSN은 객체 분할 헤드에서 클래스를 맞추는 데에 집중하여 의미론적 분할 헤드에 대해서는 충분히 다루지 않았다. 모르는 클래스에 대한 분할들을 만드는 상향식 방법을 함께 이용하면 성능이 더 나아질 것으로 보인다.

두번째로, EOPSN은 모르는 클래스를 찾기 위해 학습 데이터 위에서의 클러스터링에 의존한다. 만약, 학습 이미지들에서 보지 못한 모르는 클래스가 등장한다면모델의 성능은 상당히 저하될 것이다. 하지만, 데이터를 많이 모을수록 본 모르는 클래스들의 수가 급격하게 증가하여 제안된 모델의 장점은 극대화 될 것이다; 반면, 클래스와 무관한 물체 점수를 이용한 RPN을 이용한 베이스라인들과 같은 방법들은 아는 클래스에 대한 정답만을 사용하기 때문에 이득이 적을 것이다.

마지막으로 OPS를 위한 새로운 매트릭이 정의되어야한다. 학습된 모델이 테스트 때 실제 모르는 클래스에 대한 객체이지만 이에 대한 정답이 없는 객체들을 많이 찾을수록 실제 성능과 측정된 성능 사이의 차이는 크게 벌어질 것이다. 이를 해결하는 가장 쉬운 방법은 인간 평가를 사용하는 것이지만, 일관성에 대한 문제도 있고, 비용이 많이 든다. 이러한 문제를 완화할 수 있는 한 가지 방법은 SQ를 기반으로하는 새로운 매트릭을 만드는 것이다. 예를 들어, 모르는 클래스에 대한 확실한 거짓 양성들을 함께 고려하도록 하면 기존의 SQ보다 더 전반적인 분할 성능을 대표할 수 있게 된다.

제7장결 론

학습 때는 고려가 되지 않지만 테스트에 사용되는 모르는 클래스들이 있는 열린 집합 총체적 분할(open-set panoptic segmentation, OPS)이라는 새로운 태스크를 제시하였다. 이 태스크는 아는 클래스와 모르는 클래스 모두에 대하여 정확한 총체적 분할 지도를를 만드는 것을 목적으로 한다. 새로운 벤치마크 데이터셋을 COCO [22]를 이용하여 제시하였다.

이 어려운 태스크를 해결하기 위해 열린집합 총체적 분할 네트워크(exemplarbased open-set panoptic segmentation network, EOPSN)이라는 강한 베이스라인 모델을 제시하였다. 이 프레임워크는 우선 이미지에서 추출된 영역제안들을 클러스터링하여 새로운 클래스와 그에 해당하는 본보기들을 식별하고, 학습 도중에 다른 본보기들을 찾는다. EOPSN은 Panoptic FPN [15]을 변형한 베이스라인들보다 우수한 성능을 기록하였다. EOPSN이 어떤 하향식 총체적 분할 모델과 함께 사용될 수있기 때문에 최첨단의 네트워크에 사용하면 더 우수한 성능을 기록할 것이다.

이 연구가 컴퓨터 비전 사회에 인식 태스크를 넘어선 열린집합 문제에 대한 인식을 재고하는 것을 희망한다. OPS는 조밀한 레이블링이 필요한 대규모 데이터셋 수집을 용이하게하고 까다로운 시나리오에서 보다 현실적인 작업을 처리 할 수 있게한다.

참고 문헌

- [1] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *CVPR*, 2016.
- [2] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact: Real-time instance segmentation. In *CVPR*, 2019.
- [3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *TPAMI*, 40(4):834–848, 2018.
- [4] Bowen Cheng, Maxwell D Collins, Yukun Zhu, Ting Liu, Thomas S Huang, Hartwig Adam, and Liang-Chieh Chen. Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation. In *CVPR*, 2020.
- [5] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.
- [6] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *TIT*, 13(1):21–27, 1967.
- [7] Akshay Dhamija, Manuel Gunther, Jonathan Ventura, and Terrance Boult. The overlooked elephant of object detection: Open set. In *WACV*, 2020.
- [8] Cheng-Yang Fu, Tamara L Berg, and Alexander C Berg. Imp: Instance mask projection for high accuracy semantic segmentation of things. In *ICCV*, 2019.
- [9] Ross Girshick. Fast r-cnn. In CVPR, 2015.

- [10] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *ICCV*, 2019.
- [11] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*, 2017.
- [12] Lei Guo, Gang Xie, Xinying Xu, and Jinchang Ren. Exemplar-supported representation for effective class-incremental learning. *IEEE Access*, 8:51276–51284, 2020.
- [13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *ICCV*, 2017.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *CVPR*, 2016.
- [15] Alexander Kirillov, Ross Girshick, Kaiming He, and Piotr Dollár. Panoptic feature pyramid networks. In *CVPR*, 2019.
- [16] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *CVPR*, 2019.
- [17] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *IJCV*, 123(1):32–73, 2017.
- [18] Justin Lazarow, Kwonjoon Lee, Kunyu Shi, and Zhuowen Tu. Learning instance occlusion for panoptic segmentation. In *CVPR*, 2020.

- [19] Qizhu Li, Anurag Arnab, and Philip HS Torr. Weakly-and semi-supervised panoptic segmentation. In *ECCV*, 2018.
- [20] Yanwei Li, Xinze Chen, Zheng Zhu, Lingxi Xie, Guan Huang, Dalong Du, and Xingang Wang. Attention-guided unified network for panoptic segmentation. In CVPR, 2019.
- [21] Tsung-Yi Lin, Piotr Dollár, Ross B Girshick, Kaiming He, Bharath Hariharan, and Serge J Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017.
- [22] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In ECCV, 2014.
- [23] Huanyu Liu, Chao Peng, Changqian Yu, Jingbo Wang, Xu Liu, Gang Yu, and Wei Jiang. An end-to-end network for panoptic segmentation. In *CVPR*, 2019.
- [24] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open world. In *CVPR*, 2019.
- [25] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
- [26] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical* statistics and probability, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [27] Douglas L Medin and Marguerite M Schaffer. Context theory of classification learning. *Psychological review*, 85(3):207, 1978.

- [28] Dimity Miller, Lachlan Nicholson, Feras Dayoub, and Niko Sünderhauf. Dropout sampling for robust object detection in open-set conditions. In *ICRA*, 2018.
- [29] Lawrence Neal, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *ECCV*, 2018.
- [30] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In ICCV, 2017.
- [31] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *ICCV*, 2015.
- [32] Robert M Nosofsky. Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115(1):39, 1986.
- [33] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In NeurIPS, 2019.
- [34] Pramuditha Perera, Vlad I Morariu, Rajiv Jain, Varun Manjunatha, Curtis Wigington, Vicente Ordonez, and Vishal M Patel. Generative-discriminative feature representations for open-set recognition. In CVPR, 2020.
- [35] Trung Pham, Vijay BG Kumar, Thanh-Toan Do, Gustavo Carneiro, and Ian Reid. Bayesian semantic instance segmentation in open set world. In *ECCV*, 2018.
- [36] Lorenzo Porzi, Samuel Rota Bulo, Aleksander Colovic, and Peter Kontschieder. Seamless scene segmentation. In *CVPR*, 2019.
- [37] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *ECCV*, 2020.

- [38] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *CVPR*, 2017.
- [39] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *CVPR*, 2017.
- [40] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015.
- [41] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 115(3):211–252, 2015.
- [42] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *TPAMI*, 35(7):1757–1772, 2012.
- [43] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. Probability models for open set recognition. *TPAMI*, 36(11):2317–2324, 2014.
- [44] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *NeurIPS*, 2017.
- [45] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In CVPR, 2015.
- [46] Huiyu Wang, Yukun Zhu, Bradley Green, Hartwig Adam, Alan Yuille, and Liang-Chieh Chen. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In ECCV, 2020.

- [47] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019.
- [48] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *CVPR*, 2018.
- [49] Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. Learning semantic representations for unsupervised domain adaptation. In *ICML*, 2018.
- [50] Yuwen Xiong, Renjie Liao, Hengshuang Zhao, Rui Hu, Min Bai, Ersin Yumer, and Raquel Urtasun. Upsnet: A unified panoptic segmentation network. In CVPR, 2019.
- [51] Tien-Ju Yang, Maxwell D Collins, Yukun Zhu, Jyh-Jing Hwang, Ting Liu, Xiao Zhang, Vivienne Sze, George Papandreou, and Liang-Chieh Chen. Deeperlab: Single-shot image parser. arXiv preprint arXiv:1902.05093, 2019.
- [52] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *CVPR*, 2017.

Abstract

Exemplar-Based Open-Set

Panoptic Segmentation

Jaedong Hwang

Electrical and Computer Engineering

The Graduate School

Seoul National University

We extend panoptic segmentation to the open-world and introduce an open-set

panoptic segmentation (OPS) task. This task requires performing panoptic segmenta-

tion for not only known classes but also unknown ones that have not been acknowl-

edged during training. We investigate the practical challenges of the task and con-

struct a benchmark on top of an existing dataset, COCO. In addition, we propose a

novel exemplar-based open-set panoptic segmentation network (EOPSN) inspired by

exemplar theory. Our approach identifies a new class based on exemplars, which are

identified by clustering and employed as pseudo-ground-truths. The size of each class

increases by mining new exemplars based on the similarities to the existing ones as-

sociated with the class. We evaluate EOPSN on the proposed benchmark and demon-

strate the effectiveness of our proposals. The primary goal of our work is to draw the

attention of the community to the recognition in the open-world scenarios.

Keywords: open-set, panoptic segmentation, exemplar-based learning, clustering, min-

ing

Student Number: 2019-29143

35

감사의 글

먼저 부족한 저를 성심성의껏 지도해주신 한보형 교수님께 감사의 말씀을 드리고 싶습니다. 제가 무사히 석사과정을 마칠 수 있었던 것은 교수님의 세심한 지도와 따끔한 지적 그리고 따뜻한 격려 덕분이었습니다. 학위기간 동안 연구적인 측면 뿐 아니라 학자로서 갖추어야 할 태도에 대해 한보형 교수님께 정말 많은 것을 배웠습니다. 서울대학교 컴퓨터 비전 연구실에서 배웠던 연구에 대한 자세, 접근, 그리고 열정을 평생 간직하여 연구자로서의 삶을 성실하고 꿋꿋하게 살아가도록 하겠습니다. 또한, 학위 논문 심사를 위해 귀중한 시간을 내주시고 날카로운 질문과 조언으로 연구를 완성하는데 큰 도움을 주신 이경무 교수님, 장병탁 교수님께도 깊은 감사를 드립니다. 앞으로도 계속해서 노력하여 더 나은 연구자가 되겠습니다.

서울대학교 컴퓨터공학부에서 학부생으로 공부하던 시절, 저의 학부 졸업 후 진로에 대해서 성심성의껏 상담해주신 여러 교수님들께 감사드립니다. 특히, 장병탁교수님께는 2014년 여름 인턴을 한 이후로 정말 많은 도움을 받았습니다. 교수님의폭넓은 지식과 식견을 본받기 위해 더욱 매진하겠습니다. 학부 프로젝트 과목 결과를 논문으로 만드는 것을 권유해주신 신영길 교수님께도 감사드립니다. 덕분에학회에 참석하여 논문을 발표하며 학회의 분위기에 대해서 알 수 있었습니다.

Adobe Research에서 인턴을 하는 동안 세심한 지도를 해주신 이준영 박사님과 오승욱 박사님께 감사드립니다. 두 분 덕에 어떤 연구 주제가 좋은 주제인지에 대해 서 학문적인 관점 뿐만 아니라 실용적인 관점에서도 볼 수 있었습니다. 또한, 아이 디어를 어떻게 발전시키는지에 대해서도 배울 수 있었습니다.

학위 기간 중 가장 오랜 시간을 함께 한 컴퓨터비전 연구실 선배, 동료, 후배들에 게도 감사한 마음을 전합니다. 연구실을 소개해주시고 항상 본받고 싶은 현우형과 짧은 시간 함께 연구하며 처음에 어떤 실험부터 해야하는지 배울 수 있었습니다. 진희 누나와 종환이 형과 일하면서 논문 작성 및 연구에 대해서 많은 것을 배울 수 있었고, 한편으로는 저의 부족함이 여실히 드러나 너무나 죄송하고 감사하였습니

다. 또한, 함께 연구하는 과정에서 저 때문에 고생을 많이 하셨음에도 항상 차분하게 이야기를 해주셨던 서현씨와 종현이 형에게도 감사합니다. 인턴관련 일로 고민을 하고 있을 때 아낌없는 조언을 해주셨던 홍석이형, 궁금한 것이 있을 때 혹은 디스커 션이 필요할 때 많은 도움을 주신 유민이누나, 일채 형, 택근이 형, 민수 형, 성욱이 형, 진영이 누나, 지호씨께도 정말 감사합니다. 비디오 관련 연구를 할 때 많은 도움 을 주셨던 재유 형에게도 다시 한 번 감사드립니다. 먼저 졸업한 준하씨, 함께 수업 조교를 맡았던 미정씨, 저는 졸업하지만 박사까지 이어하는 입학동기 상혁이 형, 12 년 겨울부터 알게 되어 많은 수업을 함께 듣고 결국 같은 연구실로 진학한 학부 동기 승민이와 라이팅에 많은 도움을 주신 동완씨, 트래킹 장인이 되어가고 있는 민지씨, 목표를 향해 열심히 달려가고 있는 동환씨, 많은 이야기를 못해보아서 아쉬운 혜원 씨, 진규씨, 준혁씨, 명서씨 모두 정말 감사했습니다. 이경무 교수님의 지도를 받는 분들 중에도 감사한 분들이 너무나 많습니다. 지금은 졸업하고 새출발을 하고 계신 명섭이형, 장훈이형, 경식이형, 함께 졸업하는 승준이 형, 연구실 갈 때마다 항상 반 갑게 맞아주시던 성용이 형, 항상 긍정적이신 희원이 형, 가끔 헬스장에서 마주치던 영욱이 형, SR을 연구하는 상현씨, 포즈 예측을 연구하는 홍석씨, 군대에서 고생중 인 재린씨를 비롯해서 모든 분께 정말 감사합니다.

이외에도 이동근 선생님과 심윤만 선생님, 서울대학교 고고학과 김종일 교수님을 비롯한 많은 은사님들 덕에 무사히 학문의 길을 걸을 수 있었습니다. 또한, 학부시절 저를 지원해주었던 서암 윤세영 재단과 서울대학교 컴퓨터공학부, SNUPO 등에서 알게 된 모든 사람에게 감사합니다. 마지막으로 할머니, 할아버지 및 부모님, 동생을 비롯한 모든 친척 분들께 감사합니다. 가족이 든든한 버팀목이 되어준 덕분에 연구에 매진할 수 있었고, 이렇게 졸업을 할 수 있었습니다.

황재동 올림