Cost-Effective Methodologies for Instance Segmentation

2023.12.01

Data **M**ining & **Q**uality **A**nalytics Lab.

이 영 환

발표자 소개

❖ 이영환(Younghwan Lee)

- 고려대학교 일반대학원 산업경영공학과 재학
- **D**ata **M**ining & **Q**uality **A**nalytics Lab.(김성범 교수님)
- 석사과정(2023.03 ~)

❖ 연구 관심 분야

- Image Segmentation
- AI in Semiconductor Industry, Smart Factory

❖ E-mail

- lyh0105@korea.ac.kr

Contents

1. Introduction

Introduction

What is a segmentation?

❖ **Types of Segmentation**

Image Segmentation

: 이미지를 구성하는 픽셀을 여러 부분으로 나누는 과정.
객체를 식별하고 각 부분의 경계를 구분하여 이미지를 더욱 쉽게 분석하고 이해할 수 있게 함

Semantic Segmentation

: 이미지 내의 모든 픽셀에 대해 클래스 레이블 할당. 의미론적 분할.
: 클래스별 구분은 가능. 클래스 내 개별 인스턴스는 구분하지 않음

Instance Segmentation

: 이미지 내의 개별 객체에 고유 레이블을 할당하고 각 객체의 경계 식별
: 같은 클래스 내에서도 개별 객체를 구별

Panoptic Segmentation

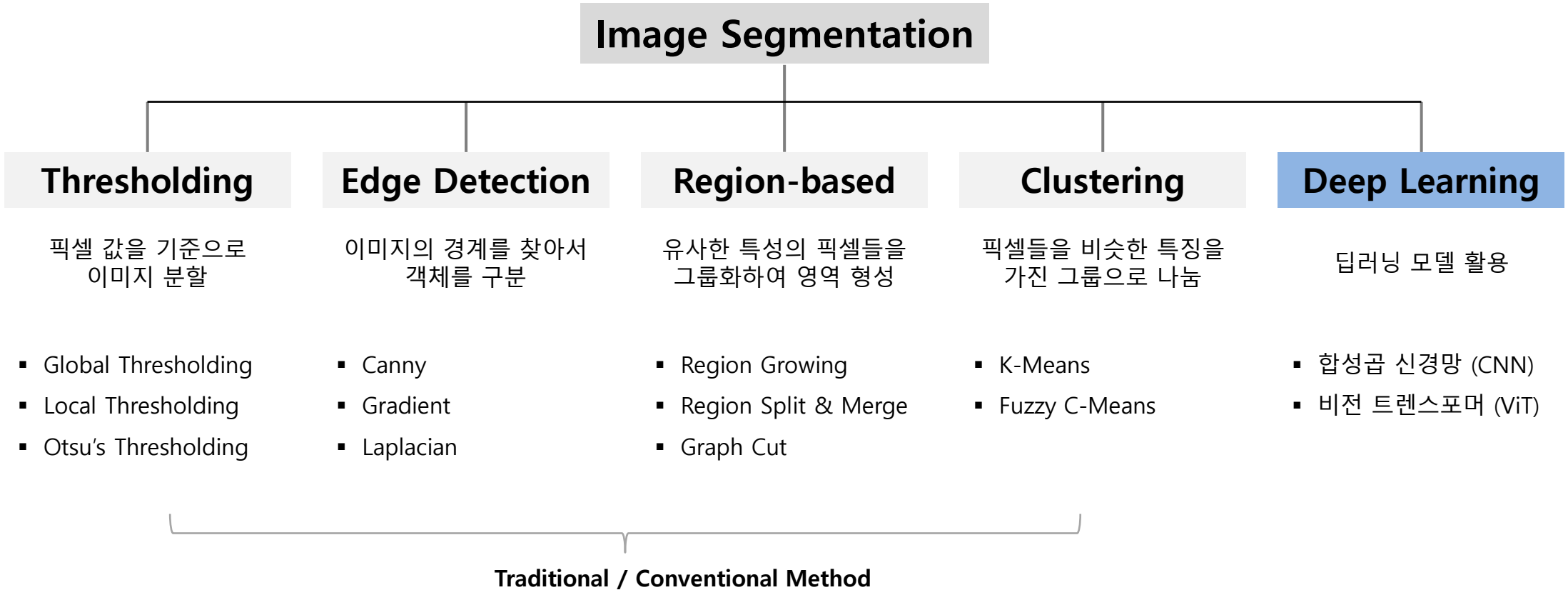
: Semantic Segmentation과 Instance Segmentation이 결합한 형태
: 의미적 구분 & 개별 객체 식별 및 세분화
: 이미지의 모든 픽셀을 클래스에 할당 / 객체 별 인스턴스도 구별

https://wiki.cloudfactory.com/docs/mp-wiki/model-families/panoptic-segmentation

Introduction

What is a segmentation?

❖ **Key techniques in Image Segmentation**



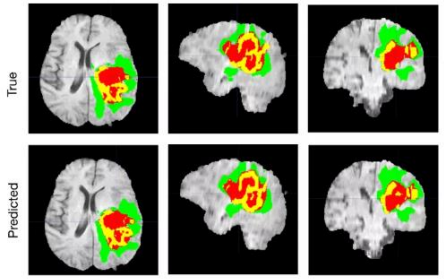
Introduction

Applications of segmentation

❖ Applications

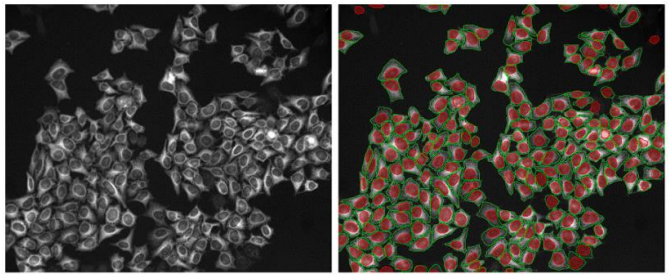
- 이미지에 대한 높은 수준의 이해가 가능하기 때문에 대표 예시인 자율주행, 의료분야(MRI, CT, X-ray) 뿐만 아니라 제조산업(불량검사), 항공, 위성, CCTV, 이미지처리, 병리학 등 그 활용 범위가 무궁무진함

[자율주행]



[의료]

[위성]



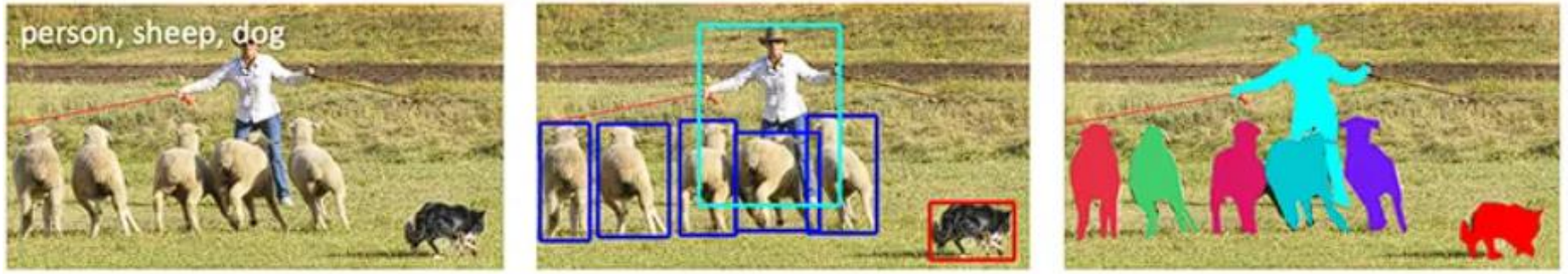
[병리학]

Introduction

Why segmentation is hard to apply?

❖ **Limitations and Challenges**

- 정답데이터를 구축하기 위해 높은 수준의 labeling cost 필요함



(a) classification (b) detection (c) segmentation

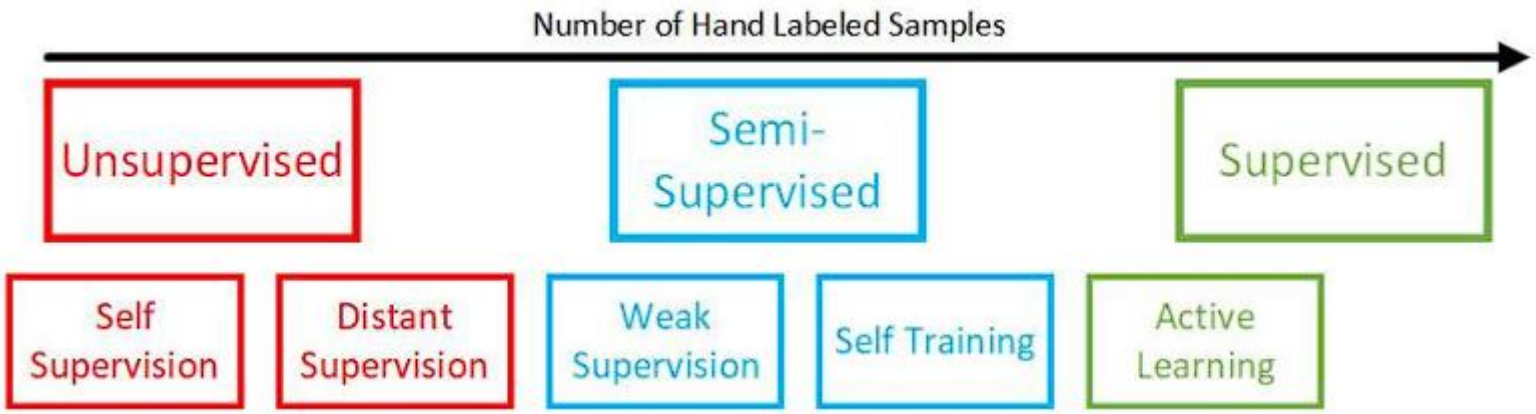
Labeling 방법 : Class 할당 < 객체를 감싸는 Bounding Box 설정 <<< 객체를 감싸는 Pixel-wise 경계 설정

https://engineering.fb.com/2016/08/25/ml-applications/segmenting-and-refining-images-with-sharpmask/

Introduction

How to solve the problem?

❖ **Approaches**



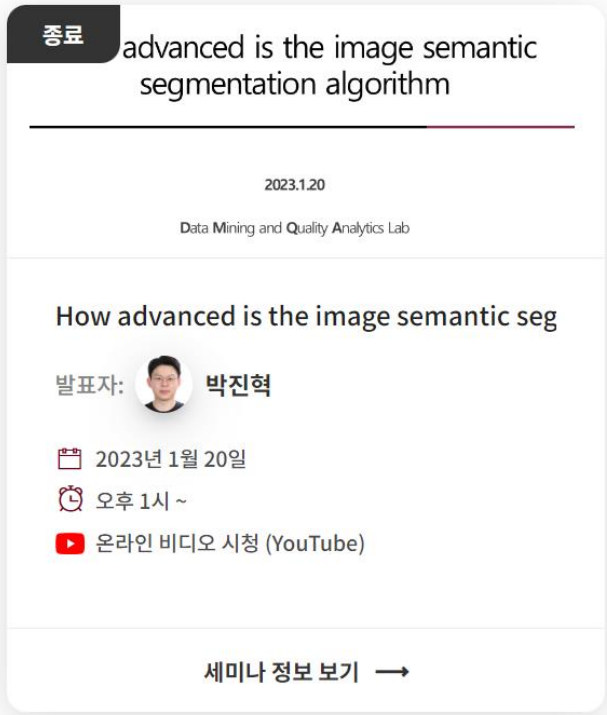
✓ **Labeling Cost를 최소화 하면서도 높은 성능을 낼 수 있는 Segmentation 방법론 필요!**

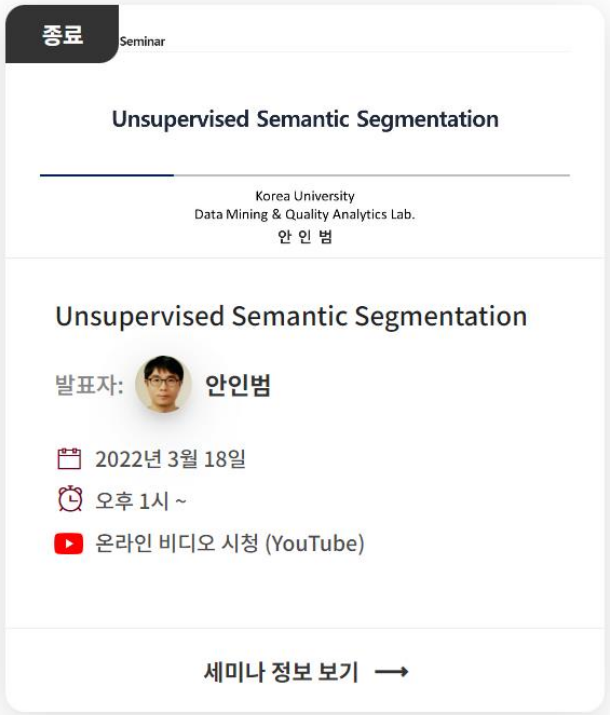
https://medium.com/@behnamsabeti/various-types-of-supervision-in-machine-learning-c7f32c190fbe

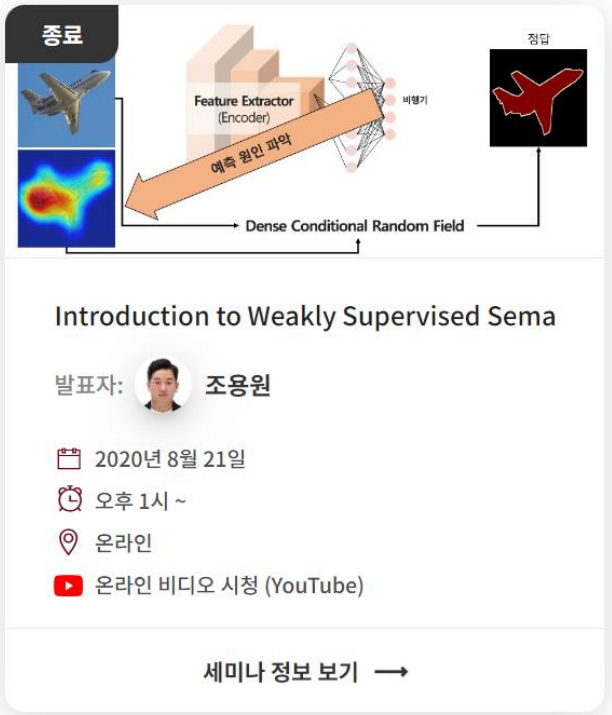
Introduction

DMQA Open Seminar

❖ **Segmentation 관련 이전 세미나 참고**







2. Paper Review

(1) CutLER : Cut and Learn for Unsupervised Object Detection and Instance Segmentation

Unsupervised Segmentation Method : CutLER

Paper

❖ **CutLER : Cut and Learn for Unsupervised Object Detection and Instance Segmentation[1]**

- 2023년에 제안된 Unsupervised Object Detection & Instance Segmentation 방법론 (CVPR, 23년 11월 기준 41회 인용)
- MaskCut, DropLoss, Self-training 을 적용한 방법론 제안

Cut and Learn for Unsupervised Object Detection and Instance Segmentation

Xudong Wang[1,2] Rohit Girdhar[1] Stella X. Yu[2,3] Ishan Misra[1]
[1]FAIR, Meta AI [2]UC Berkeley / ICSI [3]University of Michigan

Code: https://github.com/facebookresearch/CutLER

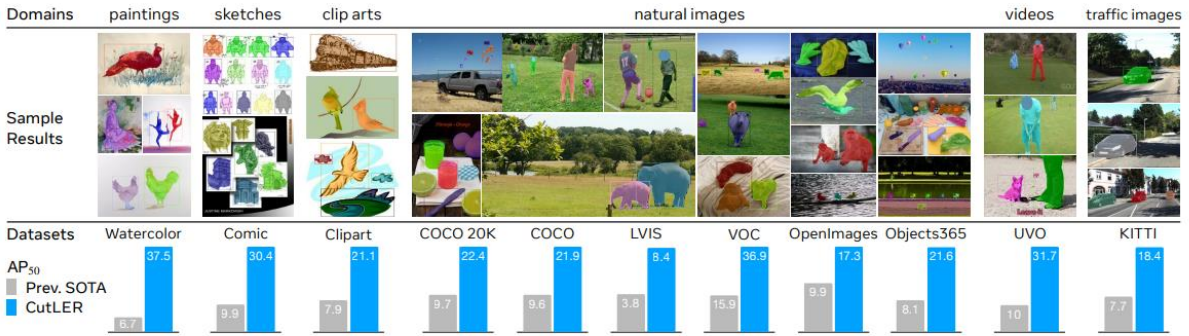


Figure 1. Zero-shot unsupervised object detection and instance segmentation using our CutLER model, which is trained without human supervision. We evaluate the model using the standard detection AP_{50}^{box}. CutLER gives a strong performance on a variety of benchmarks spanning diverse image domains - video frames, paintings, clip arts, complex scenes, *etc*. Compared to the previous state-of-the-art method, FreeSOLO [47] with a backbone of ResNet101, CutLER with a backbone of ResNet50 provides strong gains on all benchmarks, increasing performance by more than 2× on 10 of the 11 benchmarks. We evaluate [47] with its official code and checkpoint.

[1] Wang, Xudong, et al. "Cut and learn for unsupervised object detection and instance segmentation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023

Unsupervised Segmentation Method

Related Work & Motivation

❖ **선행 연구들의 한계점**

- Object Detection & Segmentation 모델 학습을 위해서는 고비용의 Annotation 필요
- 비지도학습 기반 방법론 연구 활발하지만 모든 조건을 충족시키지 못함

	DINO	LOST	TokenCut	FreeSOLO
detect multiple objects	✗	✓	✗	✓
zero-shot detector	✓	✗	✓	✗
compatible with various detection architectures	-	✓	-	✗
pretrained model for supervised detection	✓	✗	✗	✓

레이블 없이, 여러 객체를 탐지하고 분할하면서도 범용성과 효율성을 갖는 방법론 필요

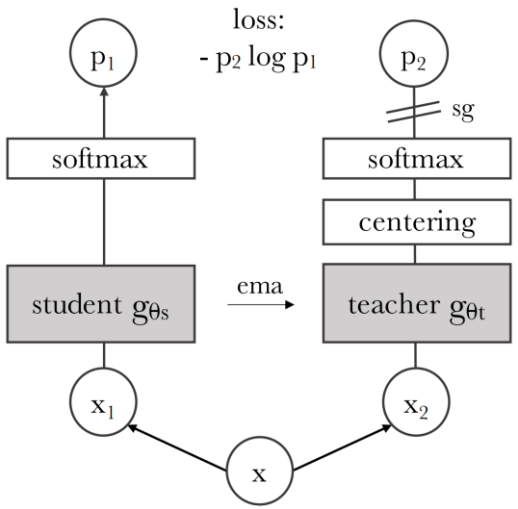
Unsupervised Segmentation Method

Related Work & Motivation

	DINO	LOST	TokenCut	FreeSOLO
detect multiple objects	✗	✓	✗	✓
zero-shot detector	✓	✗	✓	✗
compatible with various detection architectures	-	✓	-	✗
pretrained model for supervised detection	✓	✗	✗	✓

❖ **DINO : Emerging Properties in Self-Supervised Vision Transformers[3]**

- knowledge **Di**stillation (with **NO** label) 관점에서 teacher 모델이 추출한 특징을 student가 학습
- 한계점) 한 개의 object만 검출 가능. feature extractor 학습 방법(target task를 수행하진 못함)



[3] Caron, Mathilde, et al. "Emerging properties in self-supervised vision transformers." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.

Unsupervised Segmentation Method

Related Work & Motivation

❖ **NCut : Normalized Cuts and Image Segmentation[6]**

- 이미지를 전체 픽셀이 서로 연결되어 있는 그래프로 보고, 두개 또는 그 이상의 하위 그래프로 나누는 기법
- 일반화된 eigenvalue system을 해결함으로써 두 하위그래프로 그래프를 분할하는 비용을 최소화 함

▪ Generalized Eigenvalue System

$$(D - W)x = \lambda Dx$$

- D : NxN diagonal Matrix with $d_i = \sum_j W_{ij}$
- W : NXN Symmetrical Matrix

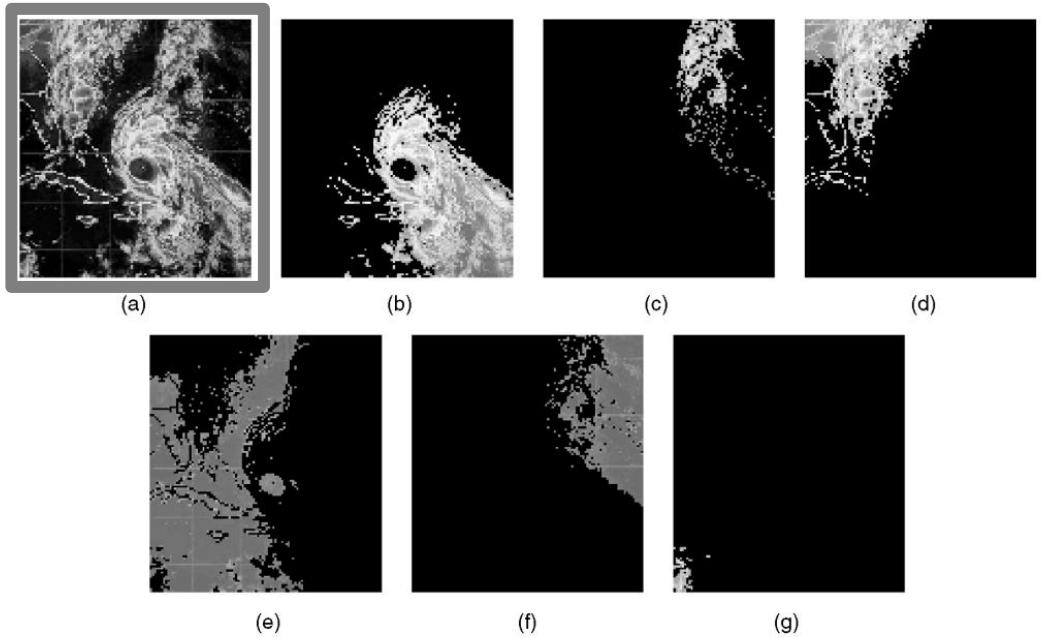


Fig. 8. (a) shows a 126×106 weather radar image. (b)-(g) show the components of the partition with $Ncut$ value less than 0.08. Parameter setting: $\sigma_I = 0.007$, $\sigma_x = 15.0$, $r = 10$.

[6] Shi, Jianbo, and Jitendra Malik. "Normalized cuts and image segmentation." *IEEE Transactions on pattern analysis and machine intelligence* 22.8 (2000)

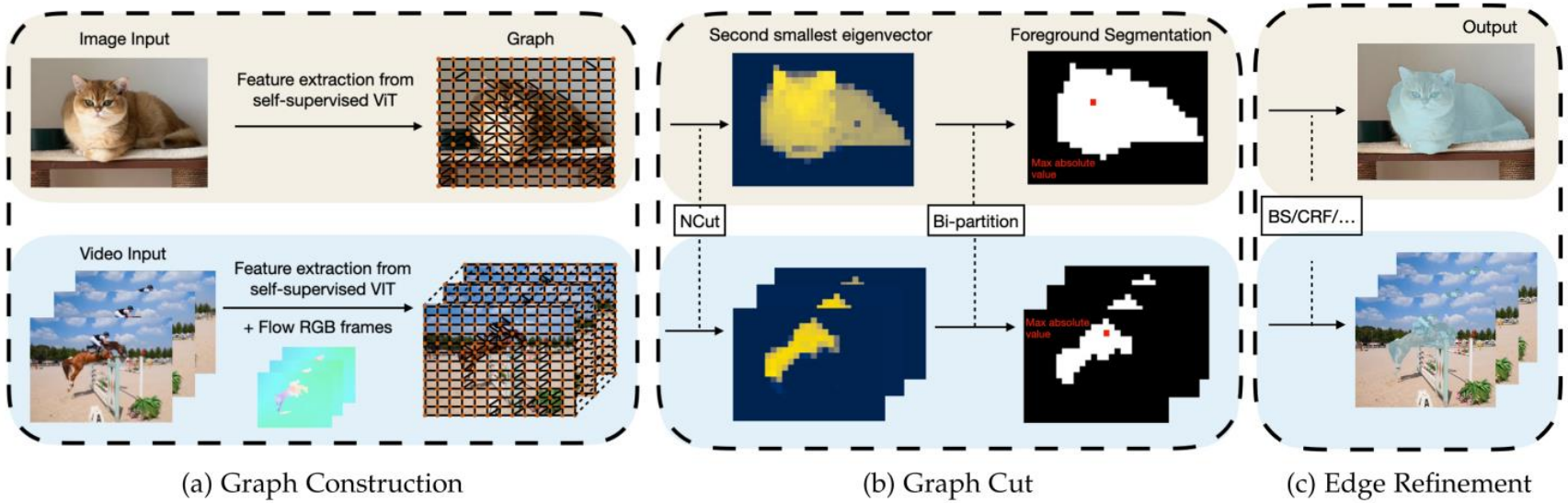
Unsupervised Segmentation Method

Related Work & Motivation

	DINO	LOST	TokenCut	FreeSOLO
detect multiple objects	✗	✓	✗	✓
zero-shot detector	✓	✗	✓	✗
compatible with various detection architectures	-	✓	-	✗
pretrained model for supervised detection	✓	✗	✗	✓

❖ **TokenCut : Segmenting Objects in Images and Videos with Self-Supervised Transformer and Normalized Cut[4]**

- 작은 패치로 자른 이미지를 DINO로 학습된 ViT 신경망에 투입하여 attention map을 얻음
- 이후 NCut(Normalized Cut)을 사용해 segmentation 수행
- 한계점) 한 개의 object만 검출 가능. pretrained model로 사용 불가능



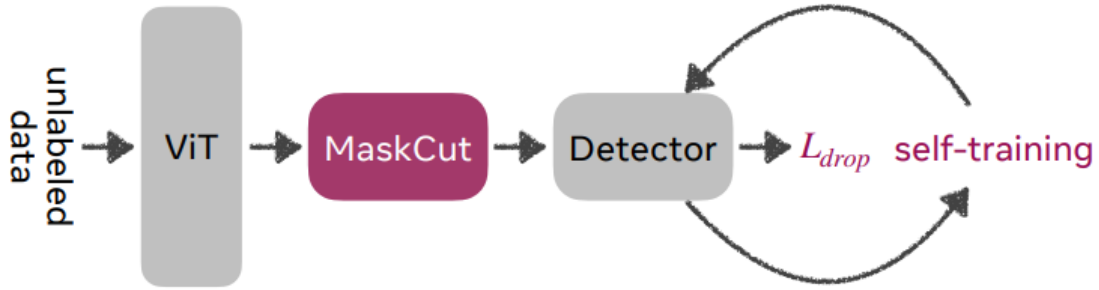
[4] Wang, Yangtao, et al. "Tokencut: Segmenting objects in images and videos with self-supervised transformer and normalized cut."
IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).

Unsupervised Segmentation Method : CutLER

Method

❖ CutLER - Overview

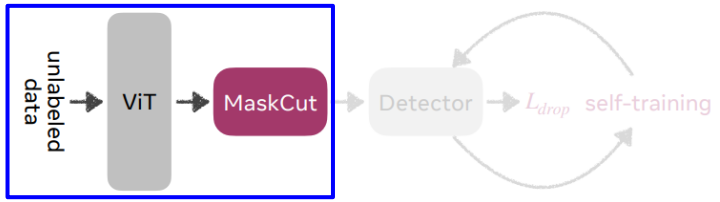
- MaskCut, Detector(w/ Droploss), Self-training 의 세가지로 구성된 단순하면서 효과적인 방법론 제시
- 기존 모델들의 한계점 극복 및 성능 향상



	DINO	LOST	TokenCut	FreeSOLO	Ours
detect multiple objects	✗	✓	✗	✓	✓
zero-shot detector	✓	✗	✓	✗	✓
compatible with various detection architectures	-	✓	-	✗	✓
pretrained model for supervised detection	✓	✗	✗	✓	✓

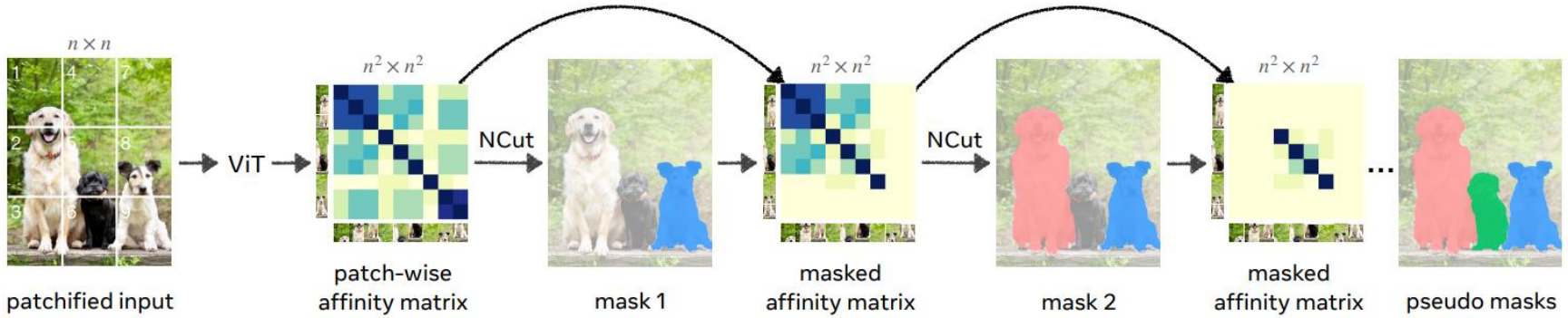
Unsupervised Segmentation Method : CutLER

Method



❖ **① MaskCut for Discovering Multiple Objects**

- 선행연구(DINO, TokenCut)의 한계점인 multi-object detection 문제를 극복하기 위해 NCut을 반복 적용
- 각 객체에 대한 Coarse Segmentation Mask 생성 목적



- NCut – Generalized Eigenvalue System

$$(D - W)x = \lambda Dx$$

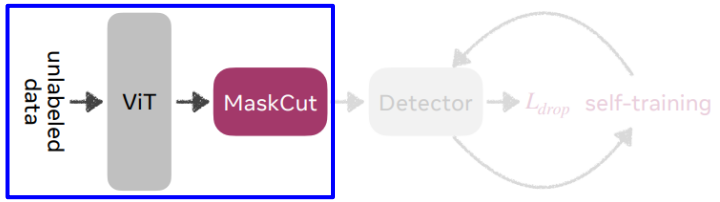
- D : NxN diagonal Matrix with $d_i = \sum_j W_{ij}$
- W : NXN Symmetrical Matrix

→ 두번째로 작은 eigenvalue λ 에 해당하는 eigenvector x 를 구함

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$$

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}$$

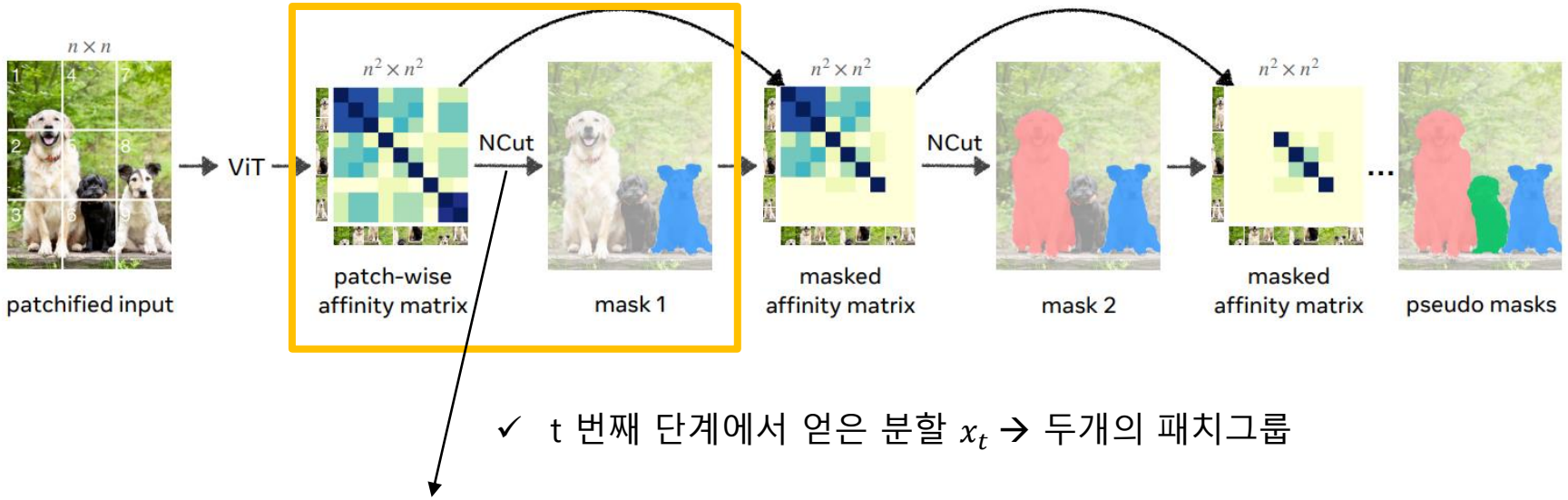
Unsupervised Segmentation Method : CutLER



Method

❖ **① MaskCut for Discovering Multiple Objects**

- 선행연구(DINO, TokenCut)의 한계점인 multi-object detection 문제를 극복하기 위해 NCut을 반복 적용
- 각 객체에 대한 Coarse Segmentation Mask 생성 목적



✓ t 번째 단계에서 얻은 분할 x_t → 두개의 패치그룹

binary mask M^t :

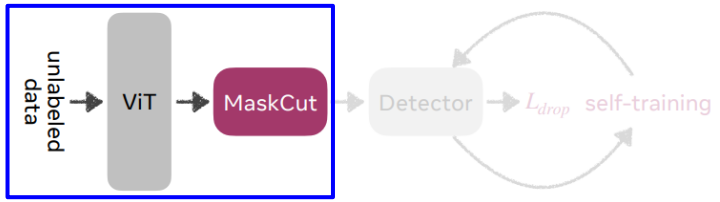
$$M_{ij}^t = \begin{cases} 1, & \text{if } M_{ij}^t \geq \text{mean}(x^t) \\ 0, & \text{otherwise.} \end{cases}$$

→ foreground

→ background

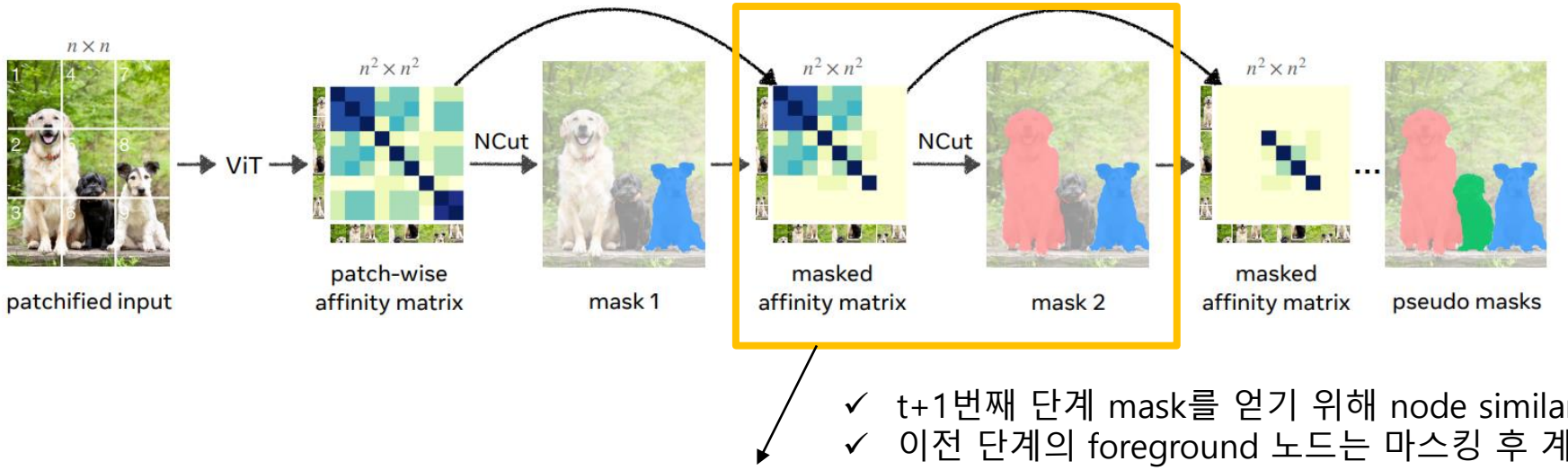
Unsupervised Segmentation Method : CutLER

Method



❖ **① MaskCut for Discovering Multiple Objects**

- 선행연구(DINO, TokenCut)의 한계점인 multi-object detection 문제를 극복하기 위해 NCut을 반복 적용
- 각 객체에 대한 Coarse Segmentation Mask 생성 목적



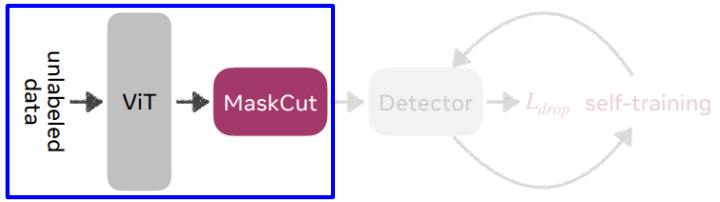
✓ t+1번째 단계 mask를 얻기 위해 node similarity 업데이트
✓ 이전 단계의 foreground 노드는 마스킹 후 계산!

Node Similarity W_{ij}^{t+1} : $$W_{ij}^{t+1} = \frac{(K_i \prod_{s=1}^{t} \hat{M}_{ij}^{s})(K_j \prod_{s=1}^{t} \hat{M}_{ij}^{s})}{\|K_i\|_2 \|K_j\|_2}$$

(※ 이전 단계의 foreground는 masking)

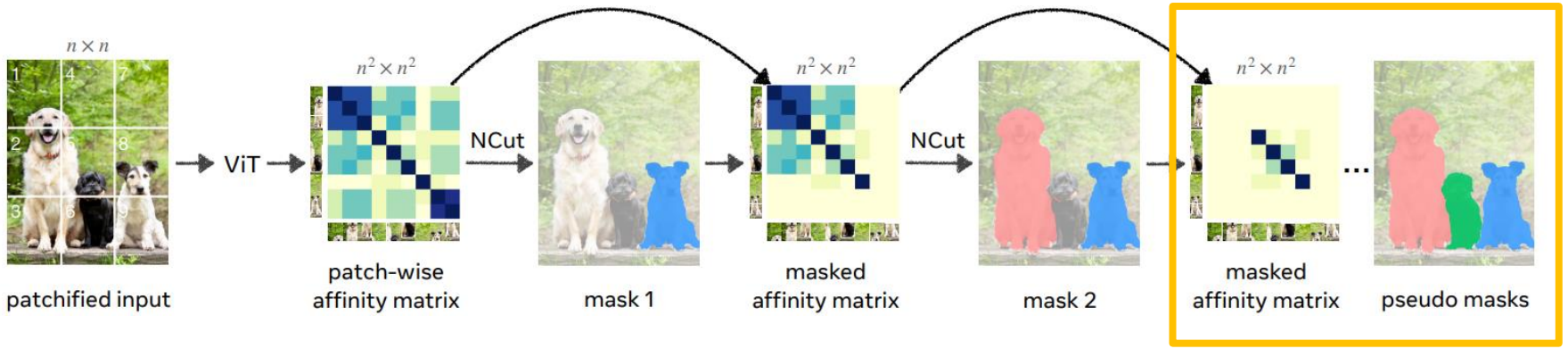
Unsupervised Segmentation Method : CutLER

Method



❖ **① MaskCut for Discovering Multiple Objects**

- 선행연구(DINO, TokenCut)의 한계점인 multi-object detection 문제를 극복하기 위해 NCut을 반복 적용
- 각 객체에 대한 Coarse Segmentation Mask 생성 목적



✓ 다중 객체를 분할하고 각 객체에 대한 이진 마스크 생성
✓ NCut의 한계점 (단일 객체 한정) 극복

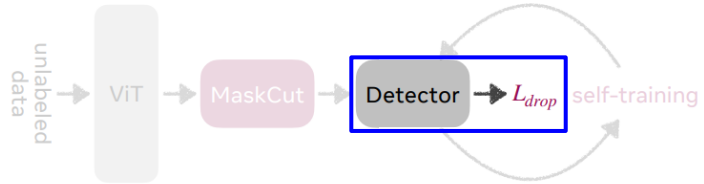
Node Similarity W_{ij}^{t+1} :

$$W_{ij}^{t+1} = \frac{(K_i \prod_{s=1}^{t} \hat{M}_{ij}^{s})(K_j \prod_{s=1}^{t} \hat{M}_{ij}^{s})}{\|K_i\|_2 \|K_j\|_2}$$

(※ 이전 단계의 foreground는 masking)

Unsupervised Segmentation Method : CutLER

Method



❖ **② DropLoss ($\mathcal{L}_{drop}$) for Exploring Image Regions**

- 기존 loss ($\mathcal{L}_{vanilla}$) : Ground-Truth와 겹치지 않는 예측영역(r_i) 에 불이익 → MaskCut이 놓친 객체는 계속 검출되지 않는 방향
- MaskCut에서 놓친 새로운 객체를 탐색할 수 있도록 DropLoss 전략 제안
- $IoU_i^{max} > \tau^{IoU}$ 인 object에 대해서만 loss 계산 (겹치는 영역이 너무 작으면 loss 계산 제외) → 새로운 객체 발견 장려

$$\mathcal{L}_{\text{drop}}(r_i) = \mathbb{1}(\text{IoU}_i^{\max} > \tau^{\text{IoU}})\mathcal{L}_{\text{vanilla}}(r_i)$$

- r_i : predicted regions
- τ^{IoU} : maximum overlap with any of the GT instances
- IoU_i^{max} : maximum IoU with GT for r^i
- $\mathcal{L}_{vanilla}$: vanilla loss function of detectors

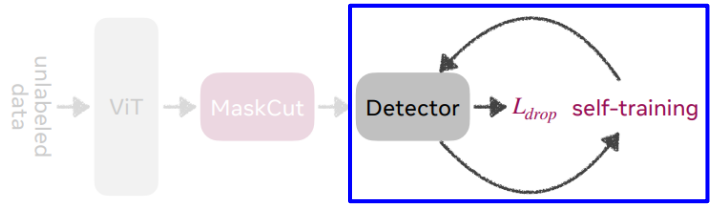
$\tau^{\text{IoU}} \rightarrow$	0	0.01	0.1	0.2
$\text{AP}_{50}^{\text{mask}}$	17.4	17.7	14.4	12.7

(d) τ^{IoU} for DropLoss.

✓ Ablation Study → τ^{IoU} = 0.01 사용

Unsupervised Segmentation Method : CutLER

Method



❖ ③ Self-Training

- 반복학습을 통한 모델 발전
- 첫번째 단계에는 MaskCut을 통해 생성된 Coarse Mask를 Pseudo Mask로 사용하여 학습
- t 단계에서 0.75-0.5t 이상의 신뢰점수를 가진 predicted mask를 → t+1 단계의 pseudo mask로 사용

✓ Self-Training 진행될 수록 pseudo mask의 질적, 양적 개선이 확인됨

	UVO			COCO		
	AP^{mask}_{50}	AP^{mask}	AP^{mask}_{75}	AP^{mask}_{50}	AP^{mask}	AP^{mask}_{75}
1 round	20.6	9.0	7.0	17.7	8.8	8.0
2 rounds	22.2	9.6	7.5	18.5	9.5	8.8
3 rounds	22.8	10.1	8.0	18.9	9.7	9.2
4 rounds	22.8	10.4	8.6	18.9	9.9	9.4

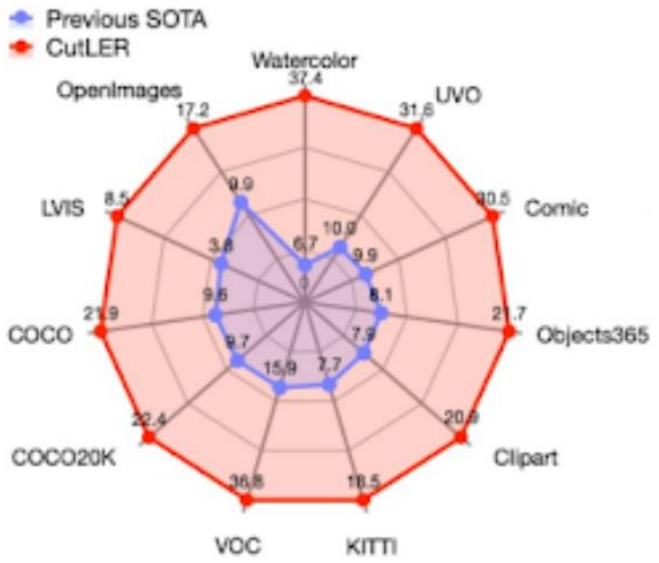
✓ Ablation Study → 반복 횟수는 3회가 적절함

Unsupervised Segmentation Method : CutLER

Experimental Results

❖ Quantitative Results

- prev. SOTA 모델(FreeSOLO) 대비 높은 수준의 성능 향상



a. zero-shot unsupervised object detection (11개의 각기 다른 데이터셋)

Datasets →	Avg.		COCO		COCO20K		VOC		LVIS		UVO		Clipart		Comic		Watercolor		KITTI		Objects365		OpenImages	
Metrics →	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR	AP_{50}	AR
Prev. SOTA [47]	9.0	13.4	9.6	12.6	9.7	12.6	15.9	21.3	3.8	6.4	10.0	14.2	7.9	15.1	9.9	16.3	6.7	16.2	7.7	7.1	8.1	10.2	9.9	14.9
CutLER	24.3	35.5	21.9	32.7	22.4	33.1	36.9	44.3	8.4	21.8	31.7	42.8	21.1	41.3	30.4	38.6	37.5	44.6	18.4	27.5	21.6	34.2	17.3	29.6
vs. prev. SOTA	+15.3	+22.1	+12.3	+20.1	+12.7	+20.5	+21.0	+23.0	+4.6	+15.4	+21.7	+28.6	+13.2	+26.2	+20.5	+22.3	+30.8	+28.4	+10.7	+20.4	+13.5	+24.0	+7.4	+14.7

b. unsupervised object detection and instance segmentation

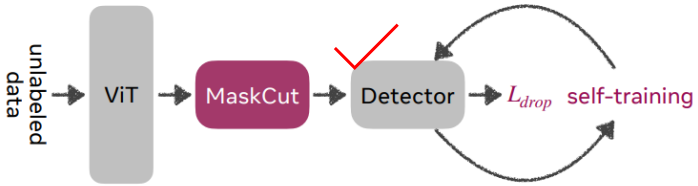
Methods	Pretrain	Detector	Init.	COCO 20K						COCO val2017					
				AP^{box}_{50}	AP^{box}_{75}	AP^{box}	AP^{mask}_{50}	AP^{mask}_{75}	AP^{mask}	AP^{box}_{50}	AP^{box}_{75}	AP^{box}	AP^{mask}_{50}	AP^{mask}_{75}	AP^{mask}
non zero-shot methods															
LOST [38]	IN+COCO	FRCNN	DINO	-	-	-	2.4	1.0	1.1	-	-	-	-	-	-
MaskDistill [42]	IN+COCO	MRCNN	MoCo	-	-	-	6.8	2.1	2.9	-	-	-	-	-	-
FreeSOLO* [47]	IN+COCO	SOLOv2	DenseCL	9.7	3.2	4.1	9.7	3.4	4.3	9.6	3.1	4.2	9.4	3.3	4.3
zero-shot methods															
DETReg [3]	IN	DDETR	SwAV	-	-	-	-	-	-	3.1	0.6	1.0	8.8	1.9	3.3
DINO [7]	IN	-	DINO	1.7	0.1	0.3	-	-	-	-	-	-	-	-	-
TokenCut [50]	IN	-	DINO	-	-	-	-	-	-	5.8	2.8	3.0	4.8	1.9	2.4
CutLER (ours)	IN	MRCNN	DINO	21.8	11.1	10.1	18.6	9.0	8.0	21.3	11.1	10.2	18.0	8.9	7.9
CutLER (ours)	IN	Cascade	DINO	22.4	12.5	11.9	19.6	10.0	9.2	21.9	11.8	12.3	18.9	9.7	9.2
vs. prev. SOTA				+12.7	+9.3	+7.8	+9.9	+6.6	+4.9	+12.3	+8.7	+8.1	+9.5	+6.4	+4.9

Unsupervised Segmentation Method : CutLER

Experimental Results

❖ Quantitative Results

- annotation data의 비율에 따른 수준 평가를 통해 pre-train model로써의 성능 확인
- label의 사용 정도(1%~100%) 에 무관하게 기존 방법론들 대비 우수한 성능 확인됨



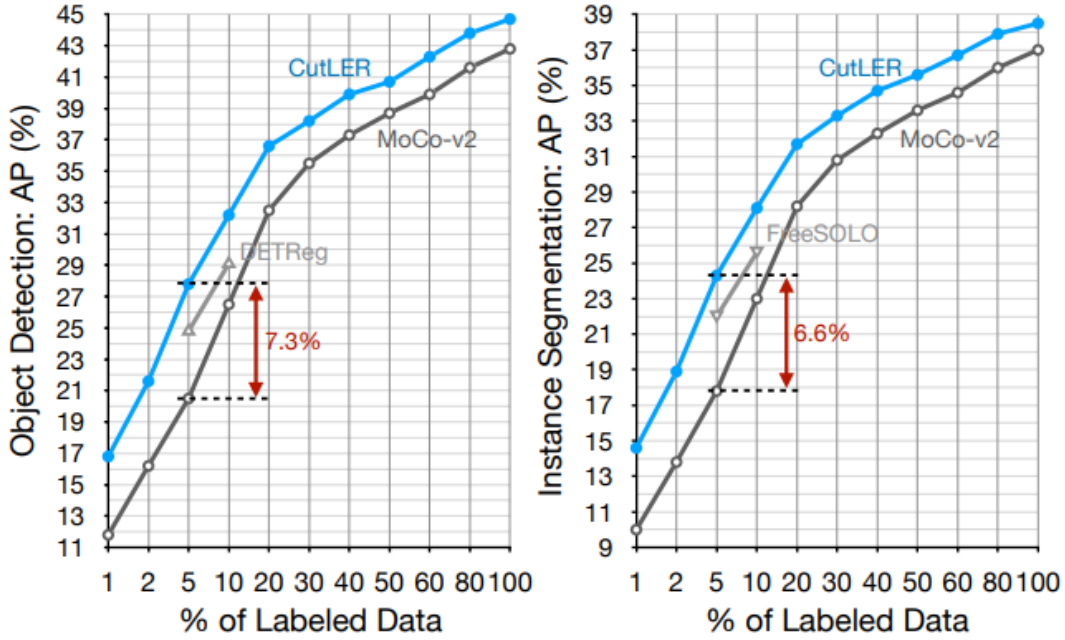


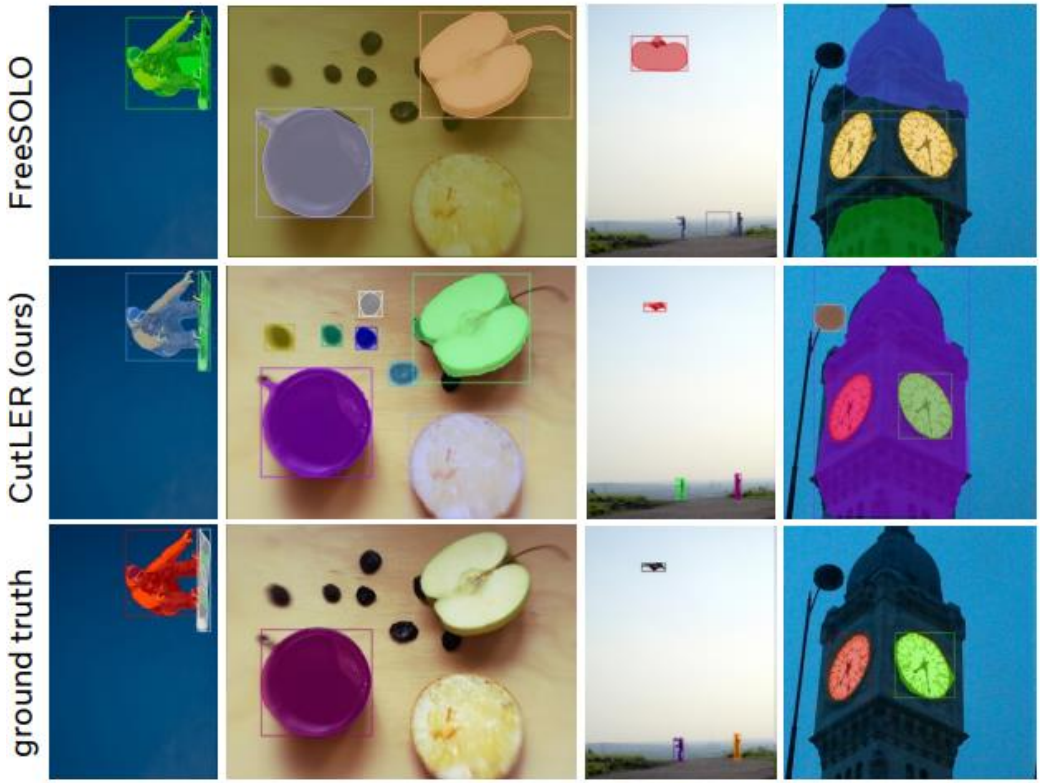
Fig 5. Finetuning CutLER for low-shot and fully supervised detection and Instance segmentation.

Unsupervised Segmentation Method : CutLER

Experimental Results

❖ Qualitative Results

- prev. SOTA 모델(FreeSOLO) 대비 더 작은 객체들을 탐지해내고 객체들을 서로 잘 분리해 냄

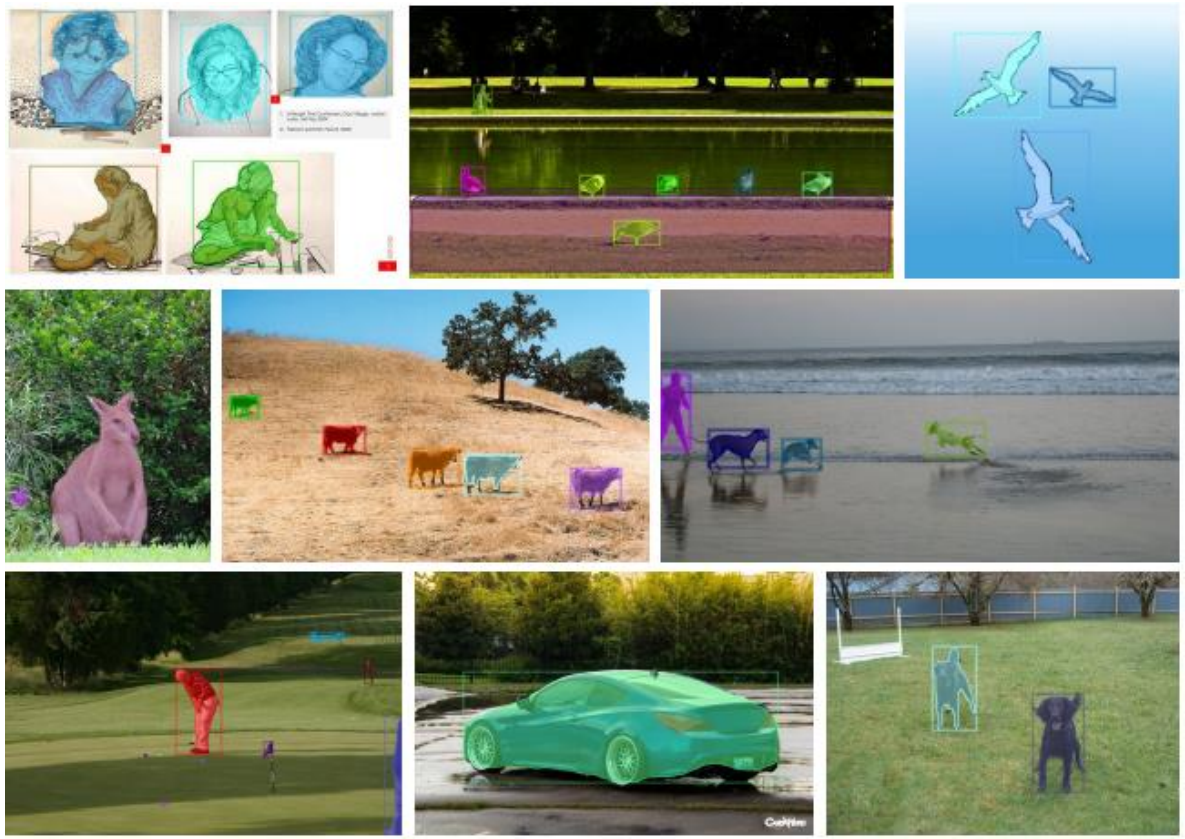


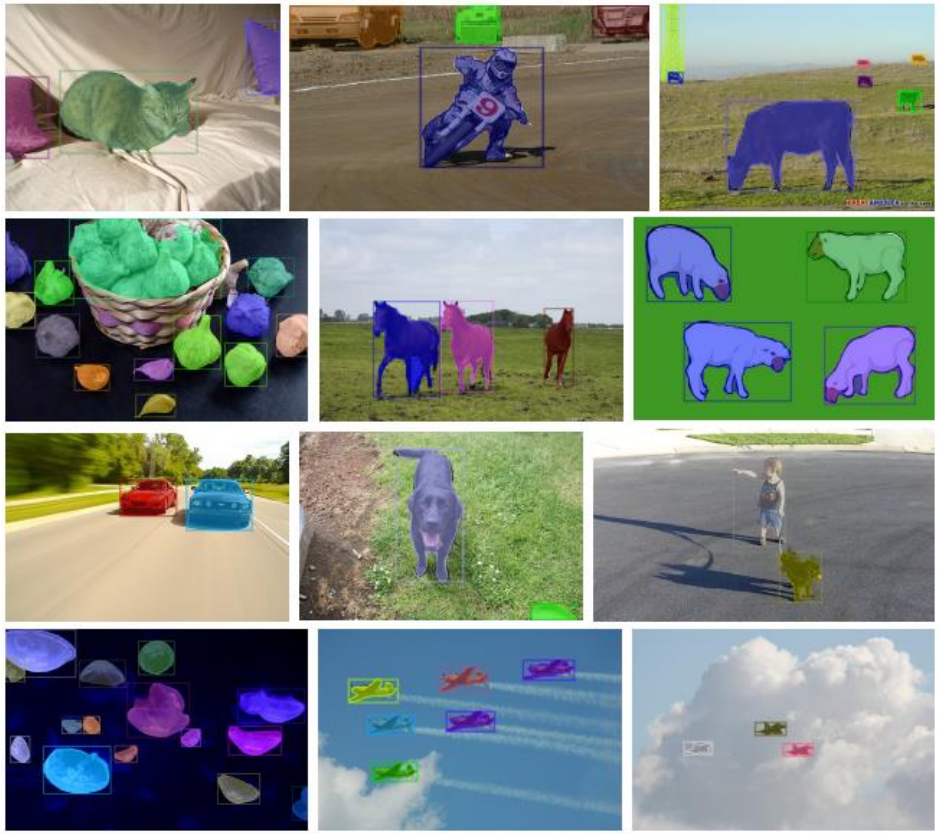
Unsupervised Segmentation Method : CutLER

Experimental Results

❖ Qualitative Results

- 개별 Instance에 대한 구분이 완벽하진 않지만 대부분의 객체에 대해 우수한 탐지 & 분할 성능을 보임





Unsupervised Segmentation Method : CutLER

Conclusion

- ✓ MaskCut, DropLoss, Self-Training 세가지 컨셉의 적용 통한 Unsupervised Object Detection & Segmentation 방법론 제안
- ✓ 레이블 없이 여러 물체에 대한 탐지 및 분할 가능
- ✓ 단순성 : 교육하기 쉽고, 다른 detection 및 segmentation 작업에 쉽게 통합 가능함
- ✓ 견고성 : 다양한 도메인에 대해 견고성을 보여줌
- ✓ 선행연구 한계점 보완 → 강력한 zero-shot 성능 확보, supervised detection을 위한 pre-train model로써의 역할 가능

	DINO	LOST	TokenCut	FreeSOLO	Ours
detect multiple objects	✗	✓	✗	✓	✓
zero-shot detector	✓	✗	✓	✗	✓
compatible with various detection architectures	-	✓	-	✗	✓
pretrained model for supervised detection	✓	✗	✗	✓	✓

2. Paper Review

(2) BoxTeacher : Exploring High-Quality Pseudo Labels for Weakly Supervised Instance Segmentation

Weakly Supervised Segmentation Method : BoxTeacher

Paper

❖ **BoxTeacher : Exploring High-Quality Pseudo Labels for Weakly Supervised Instance Segmentation[2]**

- 2023년에 제안된 Weakly Supervised Instance Segmentation 방법론 (CVPR, 23년 11월 기준 9회 인용)
- 기존의 bounding box annotation을 활용한 방식에서 더 나아가, 고품질의 pseudo mask를 사용하는 방법론 제시

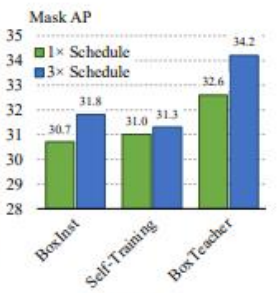
BoxTeacher: Exploring High-Quality Pseudo Labels for Weakly Supervised Instance Segmentation

Tianheng Cheng[1,*], Xinggang Wang[1], Shaoyu Chen[1,*], Qian Zhang[2], Wenyu Liu[1,†]
[1] School of EIC, Huazhong University of Science & Technology
[2] Horizon Robotics
https://github.com/hustvl/BoxTeacher

Abstract

Labeling objects with pixel-wise segmentation requires a huge amount of human labor compared to bounding boxes. Most existing methods for weakly supervised instance segmentation focus on designing heuristic losses with priors from bounding boxes. While, we find that box-supervised methods can produce some fine segmentation masks and we wonder whether the detectors could learn from these fine masks while ignoring low-quality masks. To answer this

(a)



(b)

[2] Cheng, Tianheng, et al. "Boxteacher: Exploring high-quality pseudo labels for weakly supervised instance segmentation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

Weakly Supervised Segmentation

What is Weakly Supervised Segmentation?

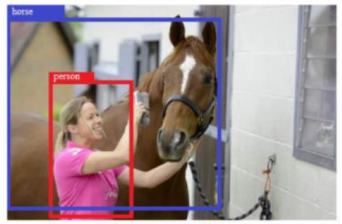
❖ Weakly Supervised Segmentation

- 지도학습 기반의 Segmentation을 위해서는 픽셀 기반의 정답이 필요함
- 픽셀기반 정답보다 약한 정보(weakly label)를 활용하여 segmentation을 진행

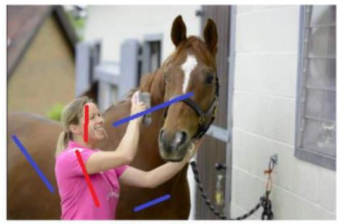
[Fully Supervised Annotation]

pixel-wise

[Weakly Supervised Annotation]



a) bounding box



b) scribble

c) point

d) image level

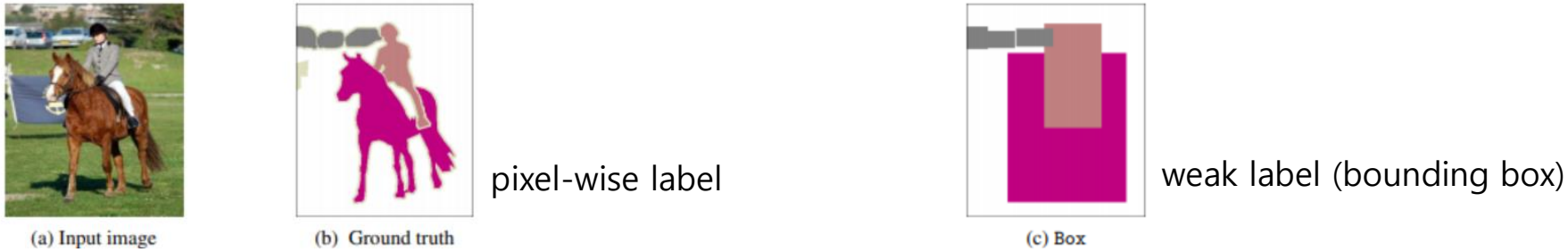
https://www.superannotate.com/blog/image-segmentation-for-machine-learning
https://velog.io/@injokim/Weakly-Supervised-Semantic-Segmentation

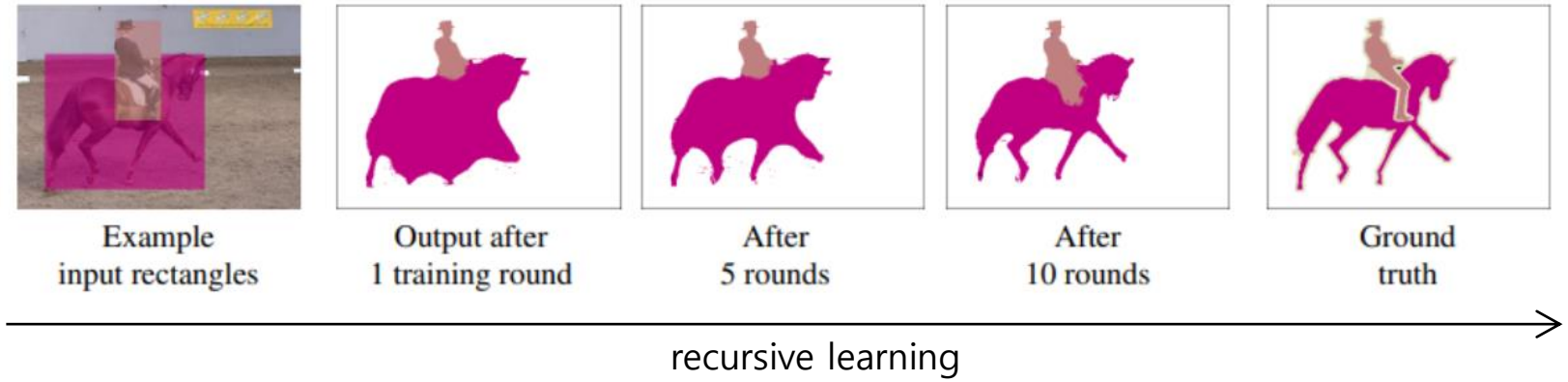
Weakly Supervised Segmentation

What is Weakly Supervised Segmentation?

❖ Weakly Supervised Segmentation

- 지도학습 기반의 Segmentation을 위해서는 픽셀 기반의 정답이 필요함
- 픽셀기반 정답보다 약한 정보(weakly label)를 활용하여 segmentation을 진행





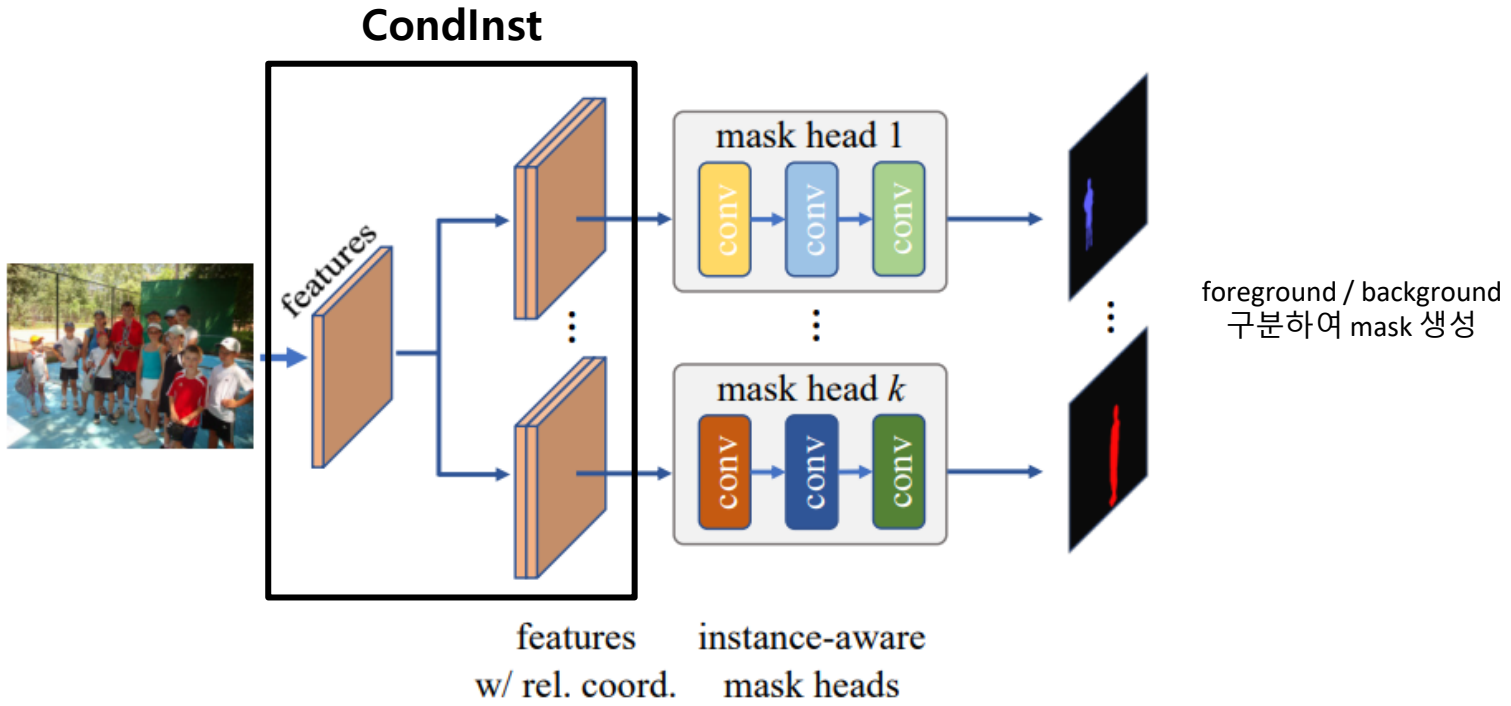
[7] Khoreva, Anna, et al. "Simple does it: Weakly supervised instance and semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Weakly Supervised Segmentation Method

Related Work & Motivation

❖ **BoxInst : High-Performance Instance Segmentation with Box Annotations[5]**

- CondInst[8] 를 사용하여 image의 instance 개수만큼의 mask head를 확보
- CondInst : RoI가 없는 Fully Convolution 방식으로 instance 세분화



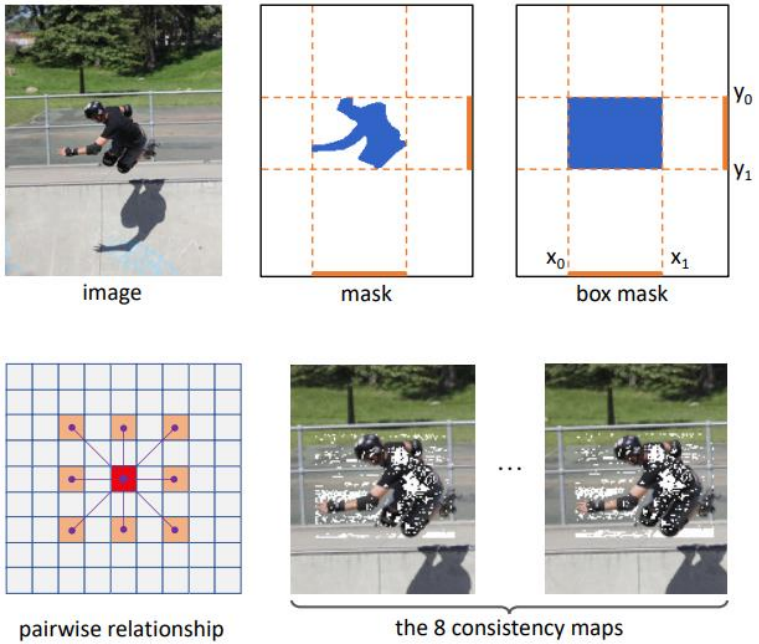
[5] Tian, Zhi, et al. "Boxinst: High-performance instance segmentation with box annotations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[8] Tian, Zhi, Chunhua Shen, and Hao Chen. "Conditional convolutions for instance segmentation." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020

Weakly Supervised Segmentation Method

Related Work & Motivation

❖ BoxInst : High-Performance Instance Segmentation with Box Annotations[5]

- Bounding Box Annotation만 활용. Instance Mask에 대한 loss를 설계하는 것에 집중



① Mask Projection Loss term for Global Localization

$$\begin{aligned} L_{proj} &= L(\text{Proj}_x(\tilde{\boldsymbol{m}}), \text{Proj}_x(\boldsymbol{b})) + L(\text{Proj}_y(\tilde{\boldsymbol{m}}), \text{Proj}_y(\boldsymbol{b})) \\ &= L(\max_y(\tilde{\boldsymbol{m}}), \max_y(\boldsymbol{b})) + L(\max_x(\tilde{\boldsymbol{m}}), \max_x(\boldsymbol{b})) \\ &= L(\tilde{\boldsymbol{l}}_x, \boldsymbol{l}_x) + L(\tilde{\boldsymbol{l}}_y, \boldsymbol{l}_y), \end{aligned}$$

- $L(\cdot, \cdot)$: dice loss as in CondInst
- Proj_x : projection the mask on x-axis
- Proj_y : projection the mask on y-axis
- l_x : 1-D segmentation mask on x-axis
- l_y : 1-D segmentation mask on y-axis

② Pairwise Relations Loss term for Local Boundaries

$$L_{pairwise} = -\frac{1}{N} \sum_{e \in E_{in}} y_e \log P(y_e = 1) + (1 - y_e) \log P(y_e = 0),$$

color similarity

$$S_e = S(\boldsymbol{c}_{i,j}, \boldsymbol{c}_{l,k}) = \exp\left(-\frac{||\boldsymbol{c}_{i,j} - \boldsymbol{c}_{l,k}||}{\theta}\right)$$

$$L_{pairwise} = -\frac{1}{N} \sum_{e \in E_{in}} \mathbb{1}_{\{S_e \geq \tau\}} \log P(y_e = 1)$$

- E_{in} : set of the edge containing at least one pixel in the box
- y_e : label for the edge e

- S_e : color similarity of edge e
- $c_{i,j}$, $c_{l,k}$: color vectors of the pixels (i, j) and (l, k) linked by edge
- θ : hyperparameter

[5] Tian, Zhi, et al. "Boxinst: High-performance instance segmentation with box annotations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Weakly Supervised Segmentation Method

Related Work & Motivation

❖ **BoxInst : High-Performance Instance Segmentation with Box Annotations[5]**

- Box-Supervised 방식으로 생성한 high-quality mask를 pseudo label로 활용

Prediction Ground Truth

BoxInst → high-quality segmentation masks
(accurate localization / fine boundaries)

BoxInst의 결과물을 pseudo label로 활용하여 weakly supervised instance segmentation의 성능을 좀 더 향상 시킬 수 있지 않을까?

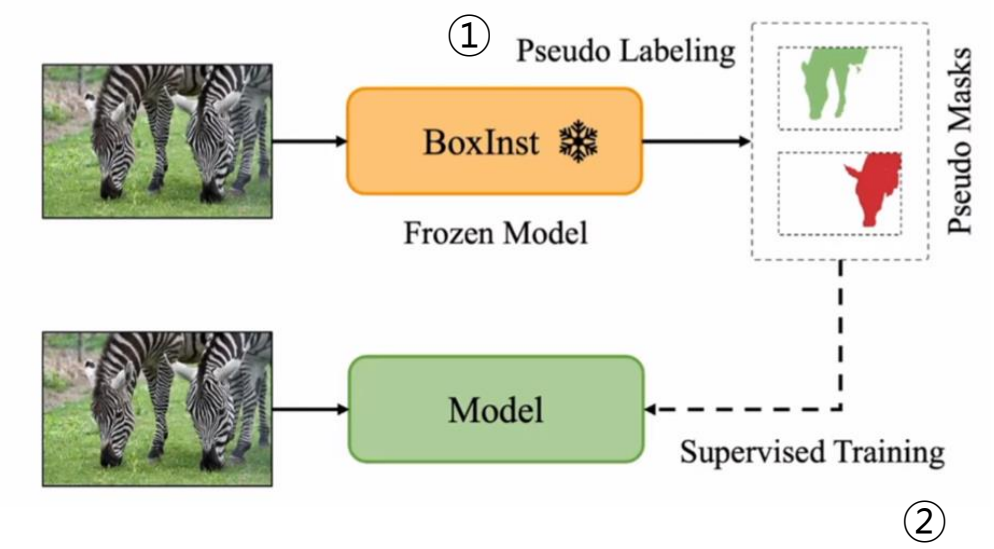
[5] Tian, Zhi, et al. "Boxinst: High-performance instance segmentation with box annotations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

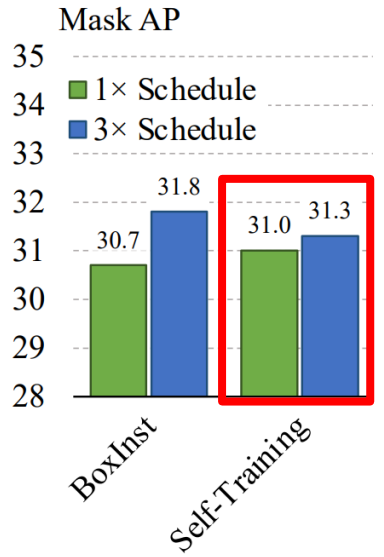
Weakly Supervised Segmentation Method

Method

❖ Naïve Self-Training with Pseudo Masks

- BoxInst를 통해 생성한 Pseudo Mask를 활용한 Self-Training





① Pre-trained BoxInst(frozen) 로 부터 pseudo mask 생성

② pseudo mask를 정답으로 하는 새로운 segmentation 모델 학습

✓ 성능 개선에 한계점 존재!

: multi-stage

: low quality mask

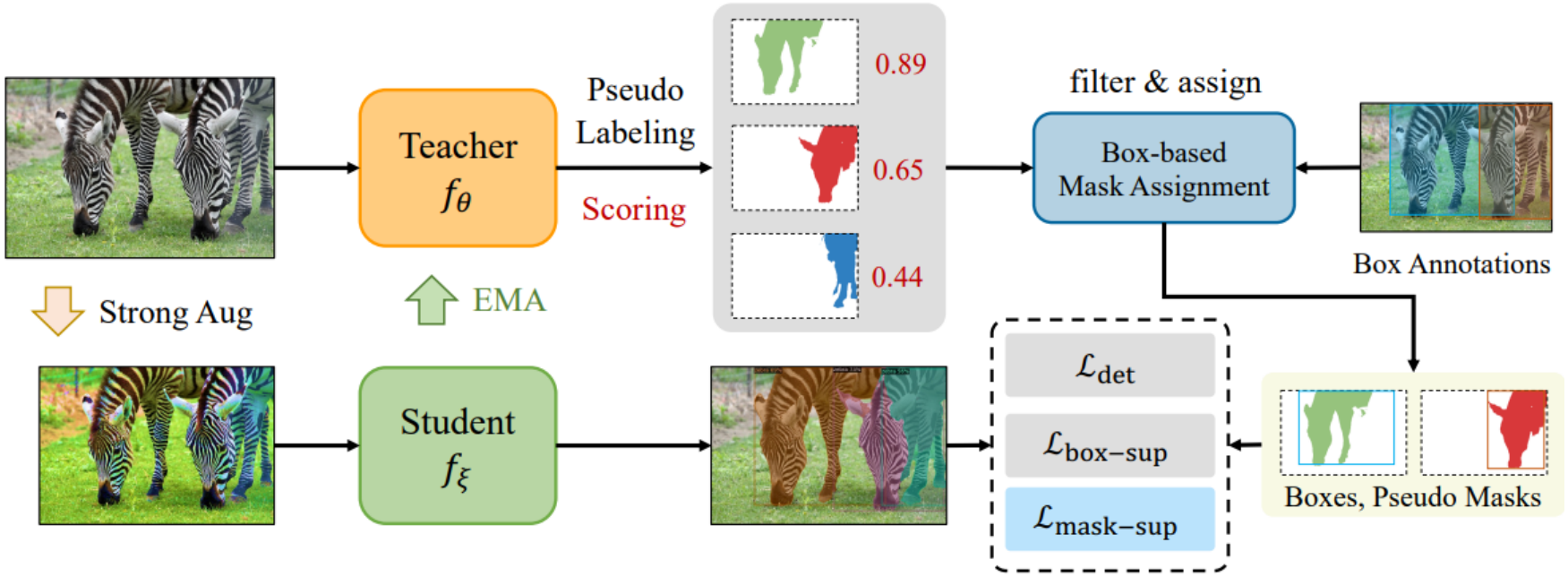
→ noisy pseudo mask

Weakly Supervised Segmentation Method : BoxTeacher

Method

❖ BoxTeacher - Overview

- Teacher가 생성한 고품질의 pseudo mask로 Student가 훈련하는 방식의 End-to-End Framework
- 선행 연구인 BoxInst와 동일한 조건에서 실험 (Box annotation만 갖는 데이터셋을 활용)

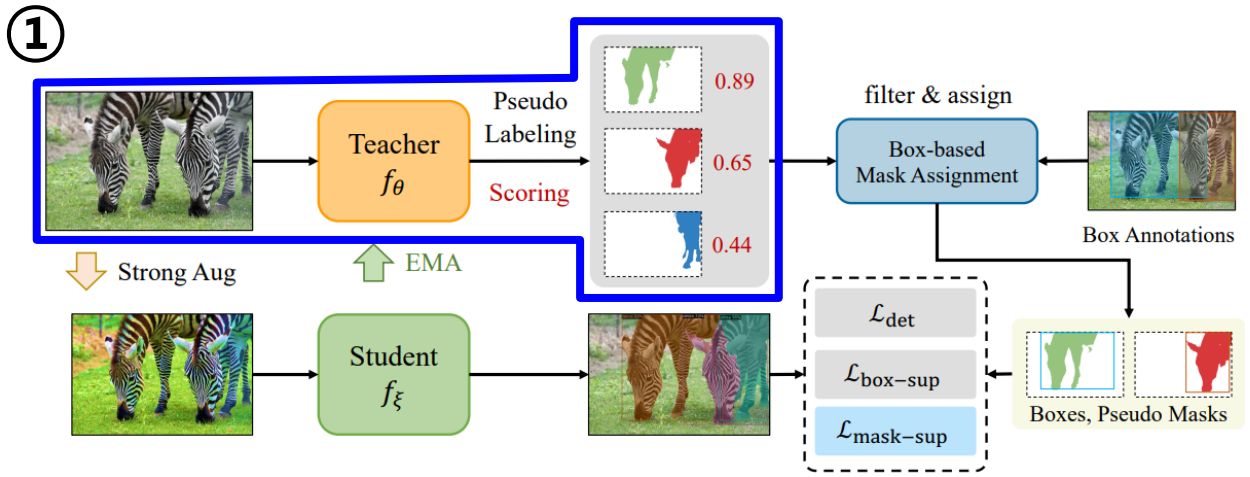


Weakly Supervised Segmentation Method : BoxTeacher

Method

❖ ① **Teacher (f_θ)**

- Teacher 모델을 활용하여 pseudo mask를 생성하고, 분류 점수와 마스크 확률에 따라 마스크 신뢰점수 추정
- 품질이 낮은 마스크는 필터링



Mask-aware Confidence Score

→ pseudo mask의 품질 추정

$$s_i = \sqrt{c_i \cdot \frac{\sum_{x,y}^{H,W} \mathbb{1}(m_{i,x,y} > \tau_m) \cdot m_{i,x,y} \cdot m^b_{i,x,y}}{\sum_{x,y}^{H,W} \mathbb{1}(m_{i,x,y} > \tau_m) \cdot m^b_{i,x,y}}}$$

threshold for binary masks (set to 0.5)

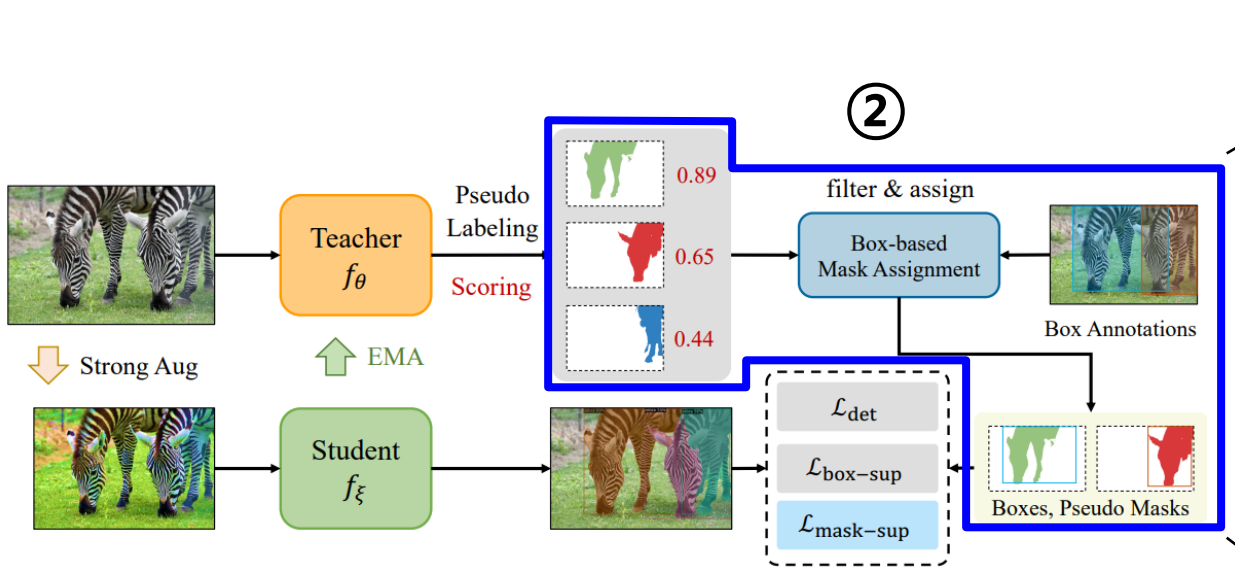
- s_i : mask-aware confidence score
- c_i : detection confidence
- m_i : sigmoid probability
- m_i^b : box-based binary mask

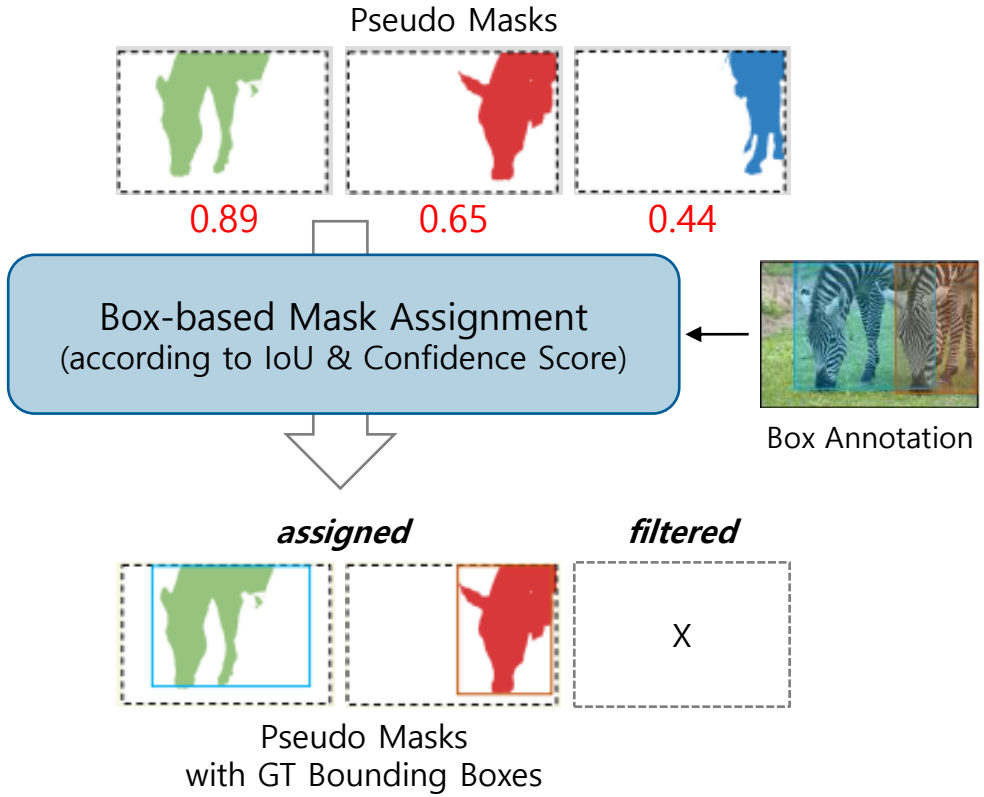
Weakly Supervised Segmentation Method : BoxTeacher

Method

❖ ② Box-based Mask Assignment

- Box-based Mask Assignment 알고리즘을 통해 pseudo mask를 GT Bounding Box에 하나씩 할당
- filter & assign의 기준은 예측과 정답사이의 IoU와 Confidence



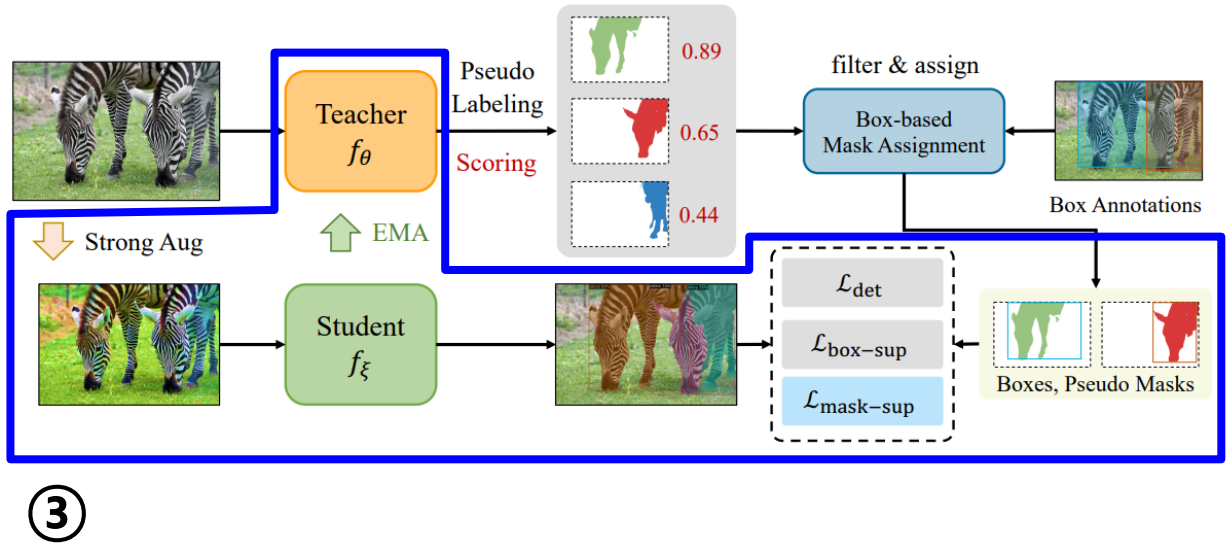


Weakly Supervised Segmentation Method : BoxTeacher

Method

❖ ③ **Student (f_ξ)**

- Student 모델은 GT box와 pseudo mask를 활용한 훈련을 통해 최적화 시키고
- EMA(Exponential Moving Average)를 통해 Teacher 모델을 업데이트 하여 고품질의 pseudo mask를 생성하게 함



③

Training Loss

$$\mathcal{L} = \mathcal{L}_{det} + \mathcal{L}_{box-sup} + \mathcal{L}_{mask-sup}$$

EMA (Exponential Moving Average)

$$f_\theta \leftarrow \alpha \cdot f_\xi + (1 - \alpha) \cdot f_\xi$$

Weakly Supervised Segmentation Method : BoxTeacher

Method

❖ ④ Training Loss

- BoxTeacher의 목표 : High-quality Mask를 활용한 fully-supervised 방식의 학습 → noisy or low-quality instance mask 제거
- pseudo mask가 가지고 있는 noise를 완화시키기 위해 pixel간 유사성을 활용하여 loss 제안 (noise-reduced mask affinity loss)

Training Loss

$$\mathcal{L} = \mathcal{L}_{det} + \mathcal{L}_{box-sup} + \mathcal{L}_{mask-sup}$$

$$\mathcal{L}_{\text{mask-sup}} = \frac{1}{N_p}\sum_{i=1}^{N_p} s_i \cdot (\lambda_p \mathcal{L}_{\text{pixel}}(m_i^p, m_i^g) + \lambda_a \mathcal{L}_{\text{affinity}}(m_i^p, m_i^g))$$

- s_i : mask-aware confidence score
- m_i^p : i-th predicted mask
- m_i^g : i-th pseudo mask
- N_p : number of valid pseudo masks

이웃 픽셀과의 refined pixel probability → local context 고려

$$\tilde{g}_i = \frac{1}{2}(g_i + \frac{1}{|\mathcal{P}|}\sum_{j\in\mathcal{P}} g_j)$$

$\mathcal{P}$: set of neighboring pixels

i번째와 j번째 pixel간의 affinity μ_{ij} → 두 픽셀간 친화도 판단

$$\mu_{ij} = \tilde{g}_i \cdot \tilde{g}_j + (1-\tilde{g}_i)\cdot(1-\tilde{g}_j)$$

noise-reduced mask affinity loss → 일관된 라벨 예측

$$\mathcal{L}_{\text{affinity}} = -\frac{\sum_{j\in\mathcal{P}} \mathbb{1}(\mu_{ij} > \tau_a)(\log(p_i \cdot p_j) + \log((1-p_i)\cdot(1-p_j)))}{\sum_{j\in\mathcal{P}} \mathbb{1}(\mu_{ij} > \tau_a)}$$

Weakly Supervised Segmentation Method : BoxTeacher

Experimental Results

❖ Quantitative Results

- Box-supervised 방법론 중 가장 우수한 성능을 보임

a. COCO Dataset

Method	Backbone	Schedule	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Mask-supervised methods.								
Mask R-CNN [23]	R-50-FPN	1×	35.5	57.0	37.8	19.5	37.6	46.0
CondInst [49]	R-50-FPN	1×	35.9	57.0	38.2	19.0	38.6	46.7
CondInst [49]	R-50-FPN	3×	37.7	58.9	40.3	20.4	40.2	48.9
CondInst [49]	R-101-FPN	3×	39.1	60.9	42.0	21.5	41.7	50.9
SOLO [54]	R-101-FPN	6×	37.8	59.5	40.4	16.4	40.6	54.2
SOLOv2 [54]	R-101-FPN	6×	39.7	60.7	42.9	17.3	42.9	57.4
Box-supervised methods.								
BoxInst [51]	R-50-FPN	3×	32.1	55.1	32.4	15.6	34.3	43.5
DiscoBox [31]	R-50-FPN	3×	32.0	53.6	32.6	11.7	33.7	48.4
BoxTeacher†	R-50-FPN	1×	32.9	54.1	34.2	17.4	36.3	43.7
BoxTeacher	R-50-FPN	3×	35.0	56.8	36.7	19.0	38.5	45.9
BBTP [25]	R-101-FPN	1×	21.1	45.5	17.2	11.2	22.0	29.8
BBAM [32]	R-101-FPN	1×	25.7	50.0	23.3	-	-	-
BoxCaseg [53]	R-101-FPN	1×	30.9	54.3	30.8	12.1	32.8	46.3
BoxInst [51]	R-101-FPN	3×	33.2	56.5	33.6	16.2	35.3	45.1
BoxLevelSet [33]	R-101-FPN	3×	33.4	56.8	34.1	15.2	36.8	46.8
BoxLevelSet [33]	R-101-DCN-FPN	3×	35.4	59.1	36.7	16.8	38.5	51.3
DiscoBox [31]	R-101-DCN-FPN	3×	35.8	59.8	36.4	16.9	38.7	52.1
BoxTeacher	R-101-FPN	3×	36.5	59.1	38.4	20.1	40.2	47.9
BoxTeacher	R-101-DCN-FPN	3×	37.6	60.3	39.7	21.0	41.8	49.3
BoxTeacher	Swin-Base-FPN	3×	40.6	65.0	42.5	23.4	44.9	54.2

b. PASCAL VOC Dataset

Method	Backbone	AP	AP_{25}	AP_{50}	AP_{70}	AP_{75}
SDI [28]	VGG-16	-	-	44.8	-	16.3
BoxInst [51]	R-50	34.3	-	59.1	-	34.2
DiscoBox [31]	R-50	-	71.4	59.8	41.7	35.5
BoxLevelSet [33]	R-50	36.3	76.3	64.2	43.9	35.9
BoxTeacher	R-50	38.6	77.6	66.4	46.1	38.7
BBTP [25]	R-101	-	75.0	58.9	30.4	21.6
Arun *et al.* [2]	R-101	-	73.1	57.7	33.5	31.2
BBAM [32]	R-101	-	76.8	63.7	39.5	31.8
BoxInst [51]	R-101	36.4	-	61.4	-	37.0
DiscoBox [31]	R-101	-	72.8	62.2	45.5	37.5
BoxLevelSet [33]	R-101	38.3	77.9	66.3	46.4	38.7
BoxTeacher	R-101	40.3	78.4	67.8	48.0	41.3

c. Cityscapes Dataset

Method	Data	AP	AP_{50}
Mask-supervised methods.			
Mask R-CNN [23]	`fine`	31.5	-
CondInst [49]	`fine`	33.0	59.3
CondInst [49]	`fine` + COCO	37.8	63.4
Box-supervised methods.			
BoxInst† [51]	`fine`	19.0	41.8
BoxInst† [51]	`fine` + COCO	24.0	51.0
BoxLevelSet† [33]	`fine`	20.7	43.3
BoxLevelSet† [33]	`fine` + COCO	22.7	46.6
BoxTeacher	`fine`	21.7	47.5
BoxTeacher	`fine` + COCO	26.8	54.2

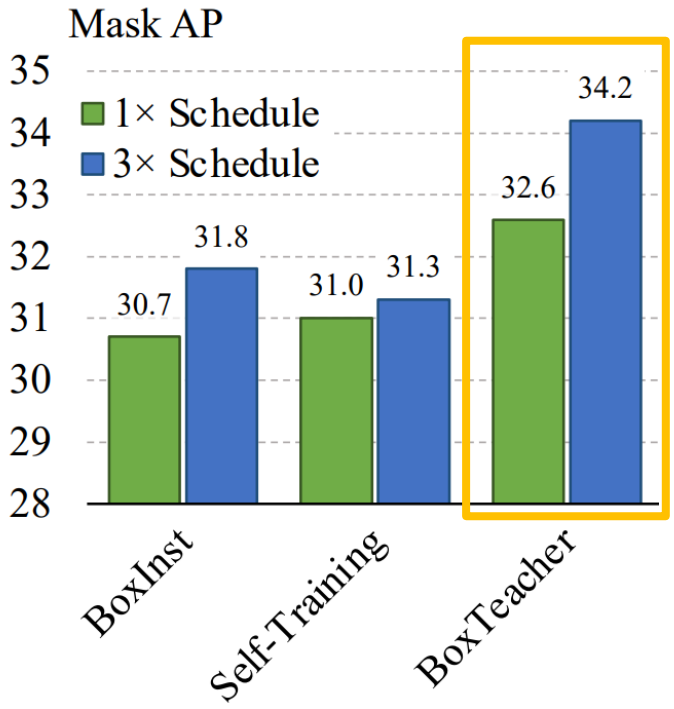
Weakly Supervised Segmentation Method : BoxTeacher

Experimental Results

❖ Quantitative Results

- BoxInst 대비 뚜렷한 개선점이 안보였던 Naïve-self training과 달리 BoxTeacher에서 높은 성능 향상 확인됨

Method	Backbone	Schedule	Pseudo Label	$AP^{\dagger}$	$AP^{\ddagger}$	AP	AP_{50}	AP_{75}
CondInst	R-50	1×	BoxInst, R-50	30.7	30.7	31.0	53.1	31.6
CondInst	R-50	3×	BoxInst, R-50	30.7	31.8	31.3	53.8	31.7
CondInst	R-50	3×	BoxInst, R-101	33.0	31.8	32.5	54.9	33.2
CondInst	R-101	3×	BoxInst, R-101	33.0	33.0	32.9	55.4	33.7
BoxTeacher	R-50	1×	End-to-End	-	-	32.6	53.5	33.8
BoxTeacher	R-50	3×	End-to-End	-	-	34.2	56.0	35.4
BoxTeacher	R-101	3×	End-to-End	-	-	35.2	57.1	36.8



Weakly Supervised Segmentation Method : BoxTeacher

Experimental Results

❖ Qualitative Results

- 복잡한 이미지에서도 세밀한 경계를 가진 high-quality segmentation 결과물을 만들어 냄

Weakly Supervised Segmentation Method : BoxTeacher

Conclusion

- ✓ 기존의 Box label을 사용한 weakly supervised 방법론의 한계점을 noisy pseudo mask로 판단
- ✓ 이를 해결하기 위해 고품질 마스크를 생성하고 훈련하는 teacher-student 모델 기반의 프레임워크인 BoxTeacher를 제안함
- ✓ mask-aware confidence score를 통해 pseudo mask의 품질을 측정하고 낮은 퀄리티의 mask를 필터링함과 동시에, 선명하고 정확한 mask를 만들어 성능 향상을 이룸
- ✓ 그렇게 생성한 pseudo mask를 활용하여 student를 optimize하기 위해 새로운 loss term을 제안함
- ✓ 다양한 데이터셋(COCO, PASCAL VOC, Cityscapes)에서 우수한 성능 확인

3. Summary

Cost-Effective Methodologies for Instance Segmentation

Conclusion

- ✓ Image Segmentation의 수요가 증가하는 반면 높은 labeling cost로 인해 접근이 어려움

- ✓ labeling cost를 최소화하면서도 우수한 성능의 Instance Segmentation을 수행하기 위한 논문 두 편 소개
  1. **CutLER** : MaskCut, DropLoss, Self-training을 활용한 Unsupervised Learning 방법론
  2. **BoxTeacher** : High-Quality Pseudo Mask를 활용한 teacher-student 기반 Weakly Supervised Learning 방법론

- ✓ 두 편 모두 '**초기 label의 고품질화**' 에 중점을 두고 있음

- ✓ Segment Anything(META AI, 2023) 의 등장으로 Segmentation 연구의 판도가 바뀌었으나
  제조, 의료 등 산업현장에서 확보되는 데이터들에 대해서는 여전히 기존 방법론들의 적절한 응용이 더 효과적일 수 있음

고맙습니다.