
How advanced is the image semantic segmentation algorithm

2023.1.20

Data Mining and Quality Analytics Lab

박진혁

Contents

1. Introduction
2. Types of semantic segmentation
3. Unsupervised semantic segmentation
4. Current unsupervised semantic segmentation trends
5. Conclusion



Introduction

❖ 발표자 소개



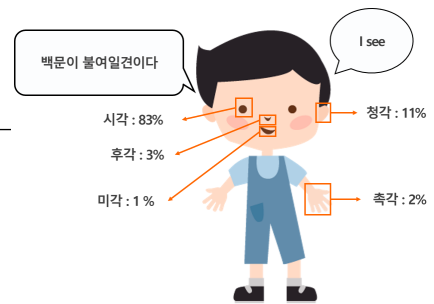
- 박진혁
- Data Mining & Quality Analytics Lab(김성범 교수님)
- 석.박사 통합과정 7학기 재학 중(2019.8 ~)

✓ 관심 분야

- Computer Vision
- Image Captioning
- Object Detection

• 박진혁

• Data Mining & Quality Analytics Lab(김성범 교수님)



❖ Computer vision이란?

- Computer vision: 사람의 시각처럼 컴퓨터도 물체나 상황을 식별, 해석할 수 있도록 하는 연구
 - Object Classification: 이미지 속 객체를 인식하여 클래스를 분류하는 연구
 - Object Detection & Localization: 이미지 또는 비디오에서 객체를 식별하는 연구
 - Object Segmentation: 이미지 및 비디오 프레임 내에서 객체를 분할하여 분류하는 연구
 - Image Captioning: 이미지의 상황을 텍스트로 설명할 수 있는 연구
 - Object tracking: 비디오 영상 내의 객체 위치를 추적하는 연구
 - Action Classification: 비디오 영상 내의 객체 행동을 인식하여 분류하는 연구



Introduction

❖ Object Segmentation

- Object Segmentation : 이미지 픽셀이 사전에 정의한 범주 중 어느 범주에 속하는지 예측
 - **Semantic segmentation**과 **Instance segmentation**으로 구성
 - **Semantic segmentation**: 같은 객체는 같은 레이블로 표현
 - **Instance segmentation**: 객체별로 다른 레이블로 표현



Introduction

❖ Object Segmentation

- Panoptic segmentation
 - Semantic segmentation과 Instance segmentation이 결합한 방식
 - 셀 수 있는 클래스(예: 사람, 동물)은 instance segmentation
 - ✓ 강아지와 고양이간의 구별
 - 셀 수 없는 클래스(예: 산, 하늘)은 semantic segmentation
 - ✓ 회색은 배경으로 셀 수 없는 클래스

Input Image



Semantic segmentation



Instance segmentation



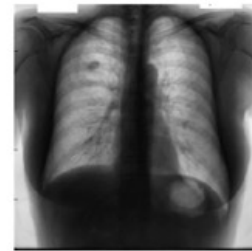
Panoptic segmentation



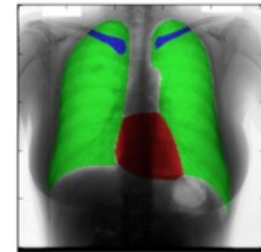
Introduction

❖ Semantic segmentation의 적용분야

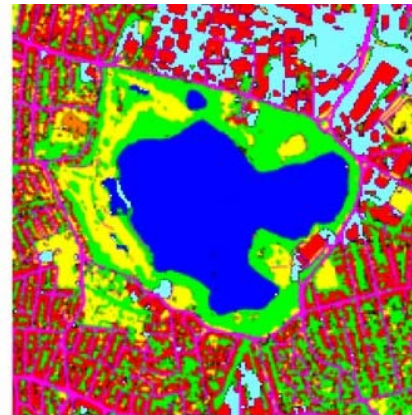
- Semantic segmentation 적용분야: 자율주행 자동차, 영상의학, 위성 이미지 분석 등 다양한 분야



Input Image



Segmented Image



Introduction



Introduction

❖ Traditional Image Segmentation

- Threshold Method
- Region Based Method
- Edge Based Method
- Clustering Based Method



Introduction

$$p(x, y) = \begin{cases} \text{white} & p(x, y) \geq T \\ \text{black} & p(x, y) < T \end{cases}$$

T : Threshold

❖ Threshold Method

- 가장 쉬운 segmentation 방법론 중 하나
- 특정 임계값(threshold)보다 큰 값을 가진 픽셀은 1 아닌 픽셀은 0으로 설정하는 방식
- 이미지는 두가지의 레이블을 가지며 클래스간 픽셀 값 차이가 큰 경우 효과적



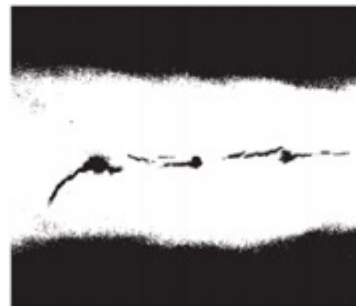
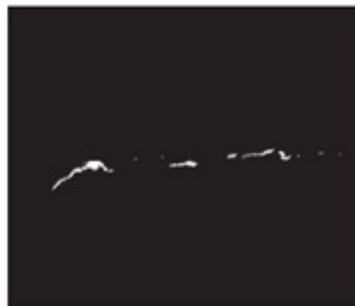
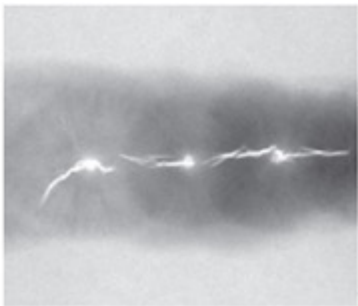
사진 출처 https://en.wikipedia.org/wiki/Thresholding_%28image_processing%29

Introduction

$$Q = \begin{cases} \text{TRUE} & \text{if the absolute difference of intensities} \\ & \text{between the seed and the pixel at } (x, y) \text{ is } \leq T \\ \text{FALSE} & \text{otherwise} \end{cases}$$

❖ Region Based Method

- 인접 픽셀간의 유사성을 찾아 공통 클래스로 그룹화하는 방법론
- 일부 픽셀을 시드 픽셀로 설정한뒤 특정 임계값을 만족하면 같은 클래스로 분류
 - 시드 픽셀과 연결된 픽셀들을 찾아 1로 레이블링
 - 각 픽셀에 정의 된 조건에 맞춰서 이미지를 생성
 - 시드 픽셀에 포함된 모든 픽셀과 이웃한 픽셀을 시드 픽셀에 추가함
 - 최종적으로 전체 이미지가 분할될 때까지 반복



Introduction

❖ Edge Based Method

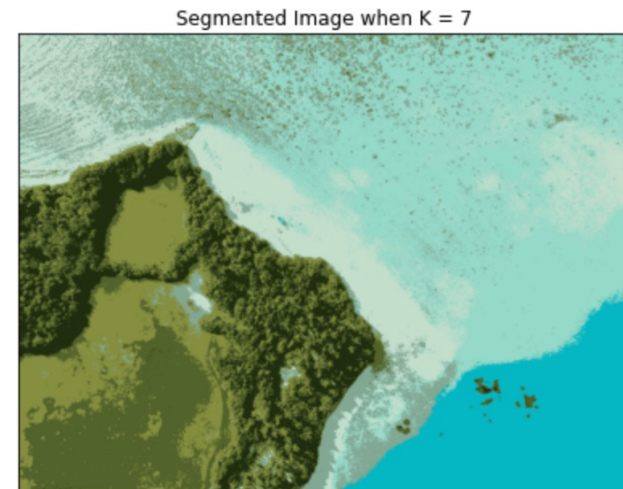
- 가장자리 정보를 사용하여 이미지안의 객체를 찾는 방법론
- 가장자리에 있는 픽셀을 찾는 과정과 찾은 픽셀들을 연결하는 과정으로 이루어짐
- 이미지에 노이즈나 가려진 부분이 존재하면 segmentation 성능이 떨어짐



Introduction

❖ Clustering Based Method

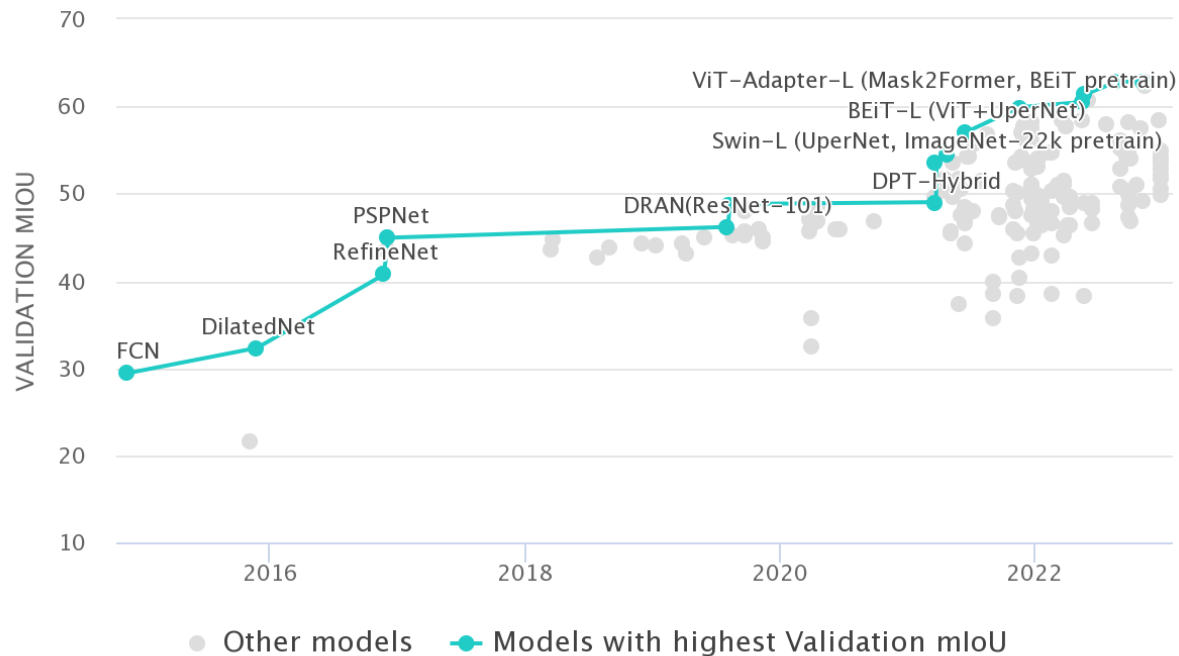
- 군집화 방법론처럼 픽셀이 어느 클래스에 해당하는지 찾아가며 segmentation 진행
 1. K개의 중심 픽셀을 임의로 지정
 2. 임의로 지정한 중심 픽셀과 가장 가까운 픽셀들을 같은 그룹에 할당
 3. 2번 결과를 바탕으로 새롭게 중심 픽셀을 지정
 4. 2~3번 과정을 중심 픽셀이 변하지 않을 때 까지 반복
- 군집 개수가 늘어날수록 더 세분화하여 클래스를 분류함



Types of semantic segmentation

❖ Semantic Segmentation의 종류

- 합성곱 신경망(Convolutional Neural Network, CNN) 기반 방법론
 - Fully convolutional network (FCN)에서 시작해서 SegFormer(2021)까지 많은 방법론 존재
- 최근에는 ViT기반의 semantic segmentation 방법론들이 SOTA를 달성함



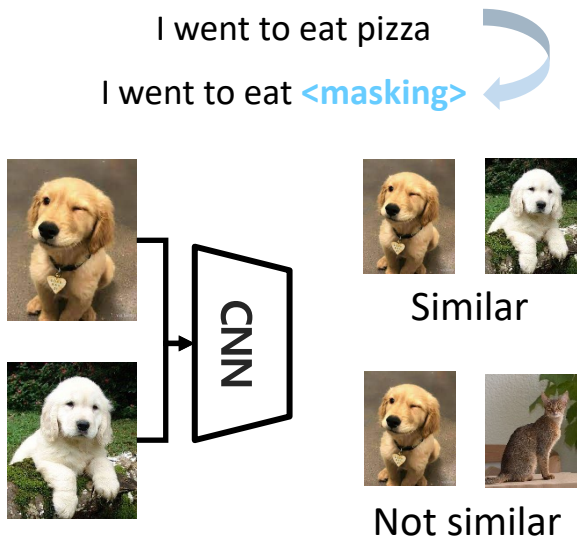
Current unsupervised semantic segmentation trends

❖ Supervised semantic segmentation의 어려움

- Semantic segmentation에서 supervised learning을 적용하기에는 많은 시간과 비용이 필요함
- Unsupervised, Semi-supervised, Weakly-supervised learning을 통해 해당 문제를 해결하고자 함

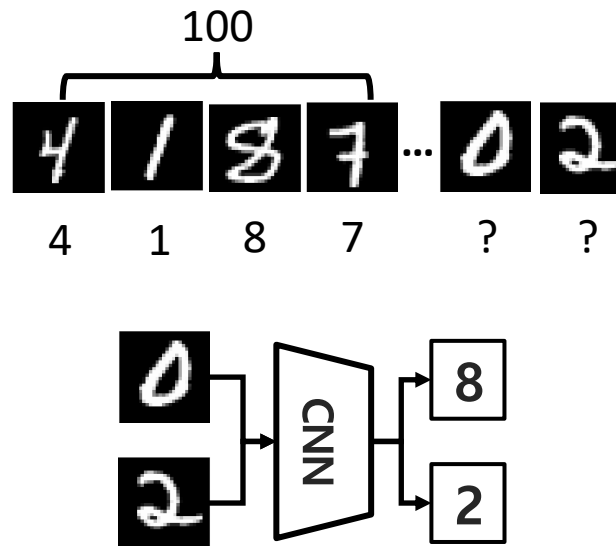
Unsupervised Learning

- 정답이 전혀 존재하지 않은 데이터를 가지고 데이터들 사이의 특징을 보다 잘 추출하여 모델 학습



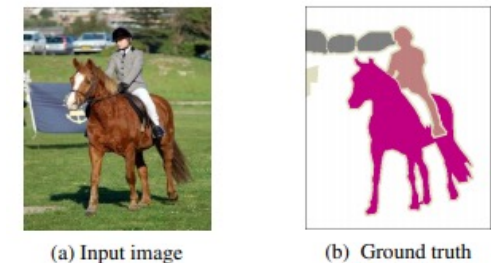
Semi-Supervised Learning

- 소수의 정답이 존재하는 데이터를 사용하여 다수의 정답이 존재하지 않는 데이터를 활용하여 모델 학습



Weakly-Supervised Learning

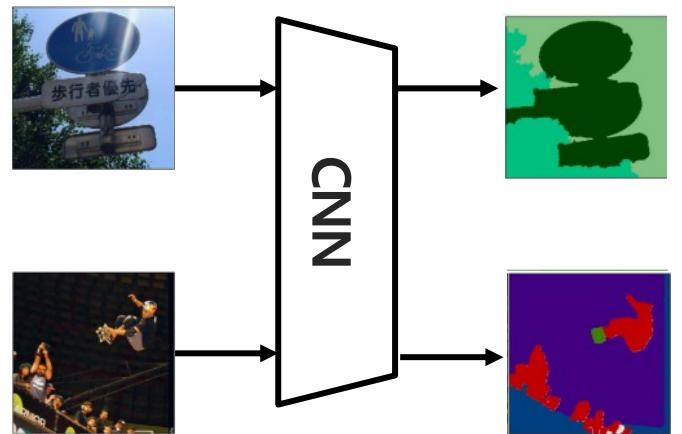
- 정보가 적은 정답데이터를 가지고 다수의 정답을 예측하는 모델 학습



Unsupervised semantic segmentation

❖ Unsupervised semantic segmentation

- Supervised semantic segmentation에서는 정답레이블 데이터셋을 구축하는데 어려움 존재
- 따라서 정답 레이블이 필요 없는 비지도 학습이 많은 딥러닝 연구에 적용됨
- Unsupervised semantic segmentation도 레이블에 존재하지 않는 데이터로 학습진행

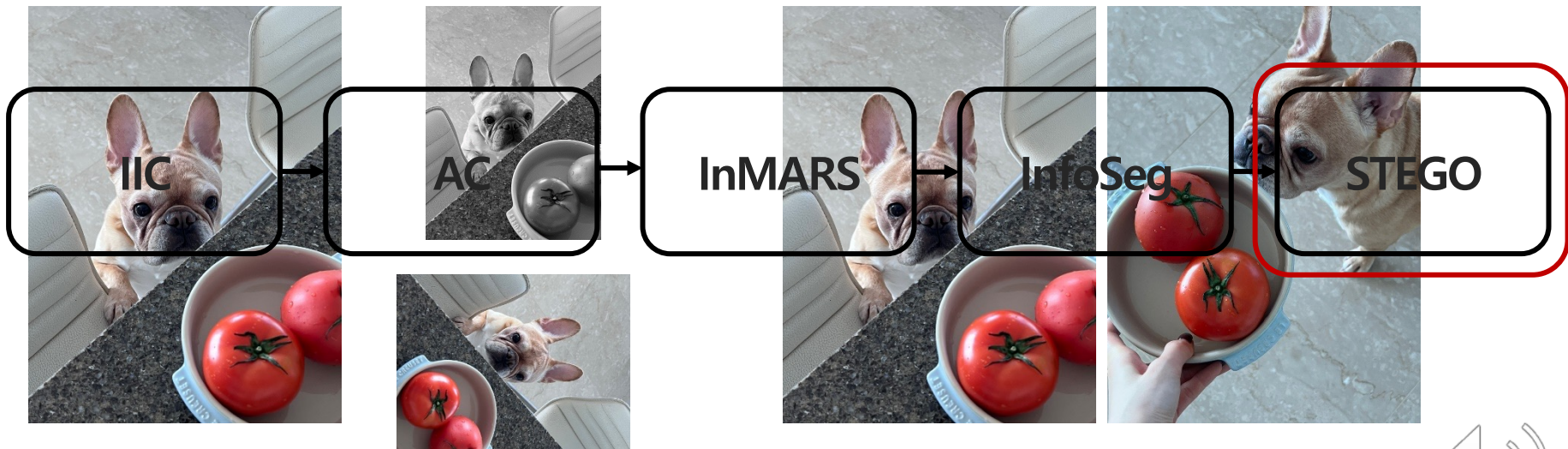
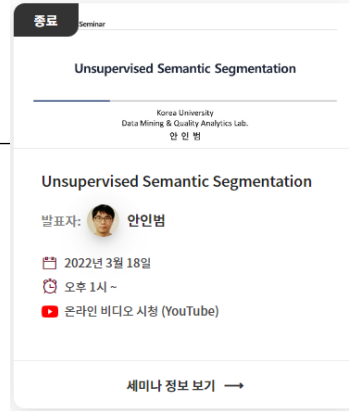


<Final output >

Current unsupervised semantic segmentation trends

❖ Unsupervised semantic segmentation

- 정답이 없는 상황에서 입력 데이터만으로 모델을 학습시킴
- 해당 연구에서는 이미지의 상호의존정보를 최대화하는 방법으로 연구진행
- 이미지의 상호의존정보를 최대화하기위해 다양한 방법론 사용
 - Augmentation, Global feature, ...



Unsupervised semantic segmentation

❖ STEGO

- 2022년 International Conference on Learning Representations(ICLR)에서 발표된 논문
- MIT, Microsoft, Cornell University, Google에서 공동연구
- Unsupervised semantic segmentation분야에서 SOTA의 성능을 보임

UNSUPERVISED SEMANTIC SEGMENTATION BY DISTILLING FEATURE CORRESPONDENCES

Mark Hamilton
MIT, Microsoft
markth@mit.edu

Zhoutong Zhang
MIT

Bharath Hariharan
Cornell University

Noah Snaveley
Cornell University, Google

William T. Freeman
MIT, Google

ABSTRACT

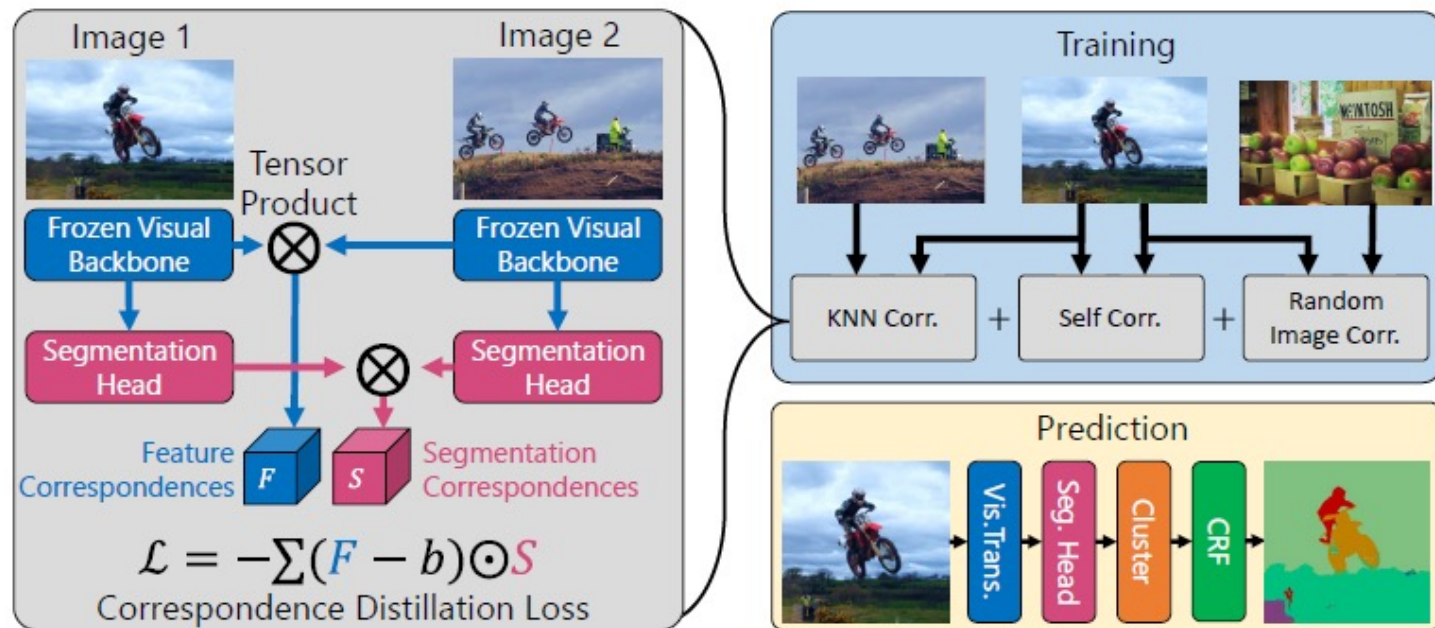
Unsupervised semantic segmentation aims to discover and localize semantically meaningful categories within image corpora without any form of annotation. To solve this task, algorithms must produce features for every pixel that are both semantically meaningful and compact enough to form distinct clusters. Unlike previous works which achieve this with a single end-to-end framework, we propose to separate feature learning from cluster compactification. Empirically, we show that current unsupervised feature learning frameworks already generate dense features whose correlations are semantically consistent. This observation motivates us to design STEGO (Self-supervised Transformer with Energy-based Graph Optimization), a novel framework that distills unsupervised features into high-quality discrete semantic labels. At the core of STEGO is a novel contrastive loss function that encourages features to form compact clusters while preserving their relationships across the corpora. STEGO yields a significant improvement over the prior state of the art, on both the CocoStuff (+14 mIoU) and Cityscapes (+9 mIoU) semantic segmentation challenges.



Unsupervised semantic segmentation

❖ Model architecture

- DINO(2021) ViT backbone을 사용하여 이미지 특징 추출
- 기존 이미지 자신, 유사한 이미지, 랜덤한 이미지를 가지고 학습에 사용
- 3가지 종류의 이미지를 통해 나온 손실함수를 합쳐서 새로운 손실함수를 정의



Unsupervised semantic segmentation

❖ Feature correspondences

- f, g : *Feature tensor*
- c : *Channel*
- (h, w) : *Spatial position(f)*
- (i, j) : *Spatial position(g)*

Image and Query Points



Self Correspondence



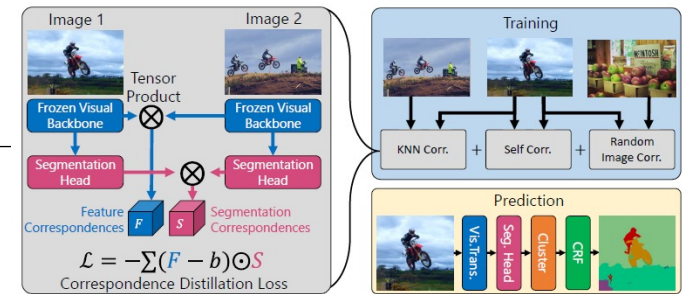
KNN Correspondence



Unsupervised semantic segmentation

❖ Model architecture

- f, g : *Feature tensor*
- c : *Channel*
- (h, w) : *Spatial position*(f)
- (i, j) : *Spatial position*(g)
- S : *Segmentation correlation tensor*
- b : *Negative pressure*(*hyper parameter*)



$$\mathcal{L}_{\text{simple-corr}}(x, y, b) := \sum_{h, w, i, j} (F_{h, w, i, j} - b) S_{h, w, i, j}$$

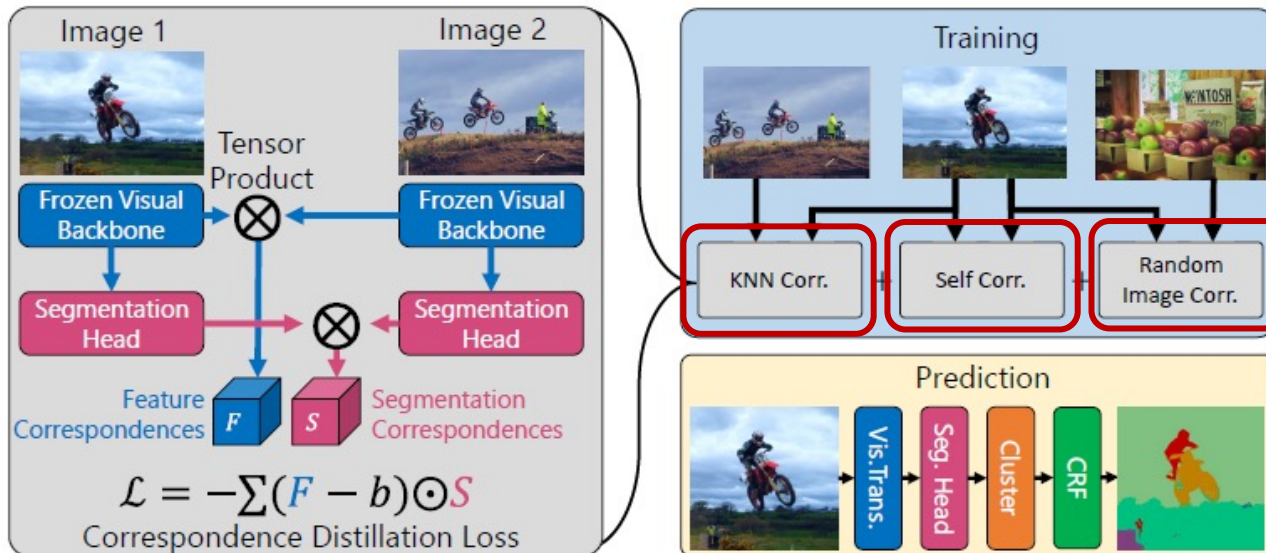
$$\mathcal{L} = \lambda_{\text{self}} \mathcal{L}_{\text{corr}}(x, x, b_{\text{self}}) + \lambda_{\text{knn}} \mathcal{L}_{\text{corr}}(x, x^{\text{knn}}, b_{\text{knn}}) + \lambda_{\text{rand}} \mathcal{L}_{\text{corr}}(x, x^{\text{rand}}, b_{\text{rand}})$$



Unsupervised semantic segmentation

❖ Model architecture

- DINO(2021) ViT backbone을 사용하여 이미지 특징 추출
- 기존 이미지 자기 자신, 유사한 이미지, 랜덤한 이미지를 가지고 학습에 사용
- 3가지 종류의 이미지를 통해 나온 손실함수를 합쳐서 새로운 손실함수를 정의



$$\mathcal{L} = \lambda_{\text{self}} \mathcal{L}_{\text{corr}}(x, x, b_{\text{self}}) + \lambda_{\text{knn}} \mathcal{L}_{\text{corr}}(x, x^{\text{knn}}, b_{\text{knn}}) + \lambda_{\text{rand}} \mathcal{L}_{\text{corr}}(x, x^{\text{rand}}, b_{\text{rand}})$$

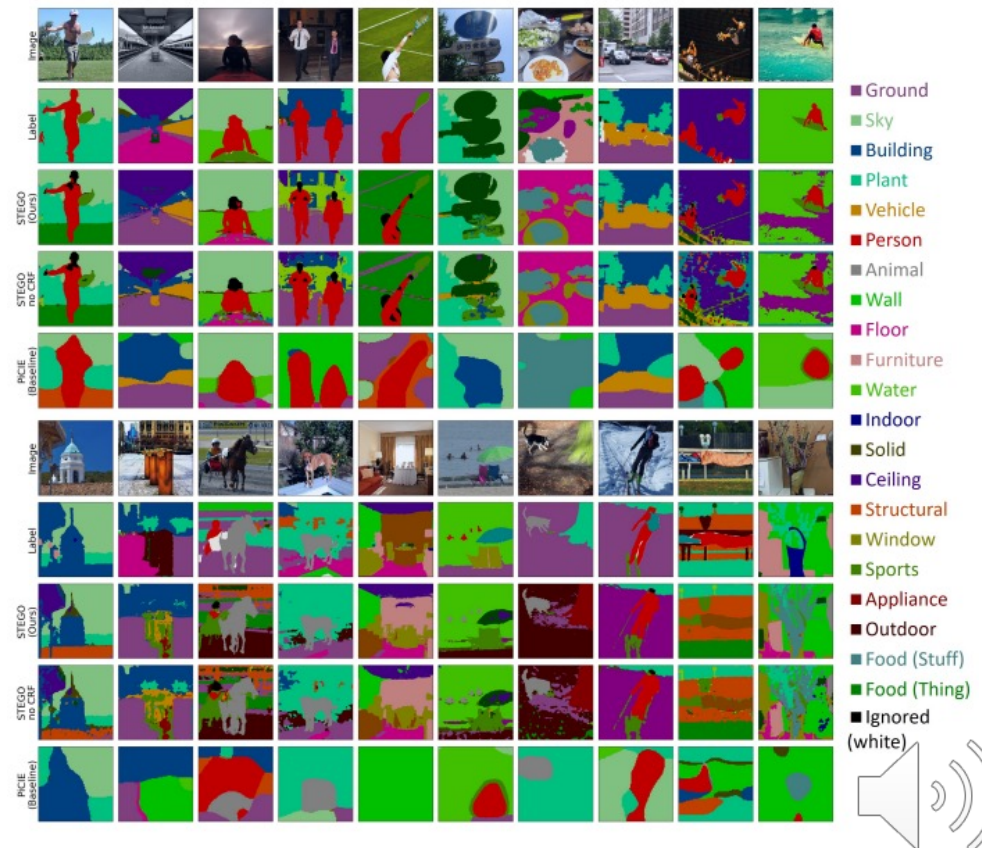
Unsupervised semantic segmentation

❖ Results

- COCOstuff(27class) 데이터셋에서 가장 좋은 성능을 보여줌
- 실제 결과를 보면 실제 레이블링보다 더 정확하게 이미지를 분할하는 경우도 존재함

Model	Unsupervised Accuracy	mIoU
ResNet50 (He et al., 2016)	24.6	8.9
MoCoV2 (Chen et al., 2020c)	25.2	10.4
DINO (Caron et al., 2021)	30.5	9.6
Deep Cluster (Caron et al., 2018)	19.9	-
SIFT (Lowe, 1999)	20.2	-
Doersch et al. (2015)	23.1	-
Isola et al. (2015)	24.3	-
AC (Ouali et al., 2020)	30.8	-
InMARS (Mirsadeghi et al., 2021)	31.0	-
IIC (Ji et al., 2019)	21.8	6.7
MDC (Cho et al., 2021)	32.2	9.8
PiCIE (Cho et al., 2021)	48.1	13.8
PiCIE + H (Cho et al., 2021)	50.0	14.4
STEGO (Ours)	56.9	28.2

<COCOstuff>

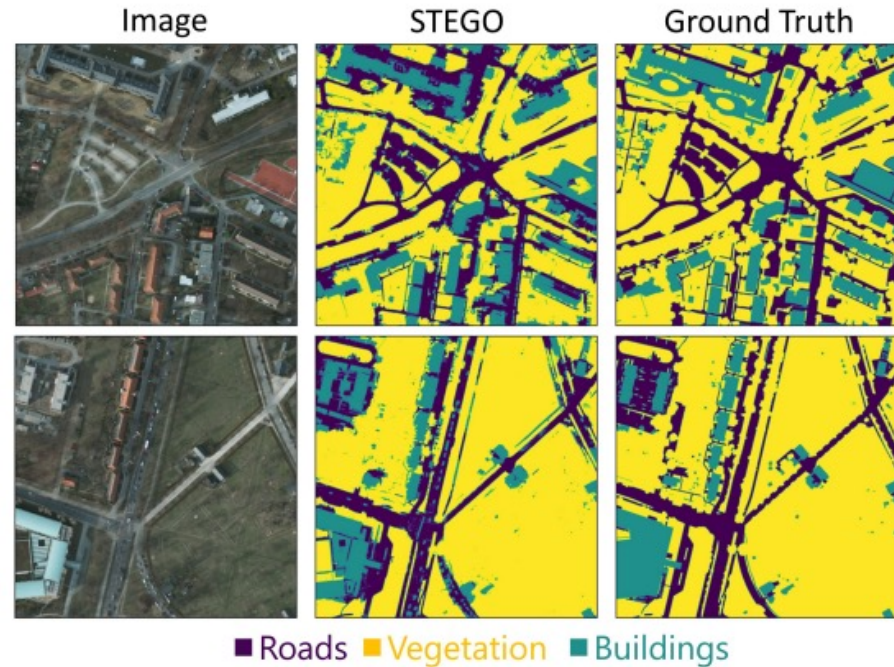


Unsupervised semantic segmentation

❖ Results

- Potsdam-3(3classes) 데이터셋에서 가장 좋은 성능을 보여줌
- 실제 결과를 보면 실제 레이블링보다 더 정확하게 이미지를 분할하는 경우도 존재함

Model	Unsup. Acc.
Random CNN (Ji et al., 2019)	38.2
K-Means (Pedregosa et al., 2011)	45.7
SIFT (Lowe, 1999)	38.2
Doersch et al. (2015)	49.6
Isola et al. (2015)	63.9
Deep Cluster (Caron et al., 2018)	41.7
IIC (Ji et al., 2019)	65.1
STEGO (Ours)	77.0



<Potsdam-3>

Conclusion

❖ Segmentation

- 이미지 픽셀이 사전에 정의한 범주 중 어느 범주에 속하는지 예측
 - Semantic segmentation: 같은 객체는 같은 레이블로 표현
 - Instance segmentation: 객체별로 다른 레이블로 표현
 - Panoptic segmentation: Semantic segmentation과 Instance segmentation이 결합한 방식

❖ Unsupervised semantic segmentation

- Unsupervised semantic segmentation은 정답 레이블이 존재하지 않는 이미지로 학습진행
- 대부분의 연구들이 상호의존정보를 최대화하는 방법론으로 진행
- 최근에는 feature correspondance를 사용한 방법론이 가장 좋은 성능을 보임(STEGO)



감사합니다

