Open DMQA Seminar

Unsupervised Semantic Segmentation

Korea University
Data Mining & Quality Analytics Lab.
안 인 범

2022. 3. 18

발표자 소개

- **안인범 (Inbum Ahn)**
  - 고려대학교 산업경영공학과 Data Mining & Quality Analysis Lab.
  - M.S. Student (2020.9 ~ Present)
  - 지도 교수 : 김성범 교수님

- **Research Interest**
  - Deep Learning for visual inspection
  - Unsupervised semantic segmentation

- **Contact**
  - E-mail : ahninb20@korea.ac.kr

Contents

I. Introduction

II. Unsupervised semantic segmentation

III. Mutual information maximization

IV. Methods

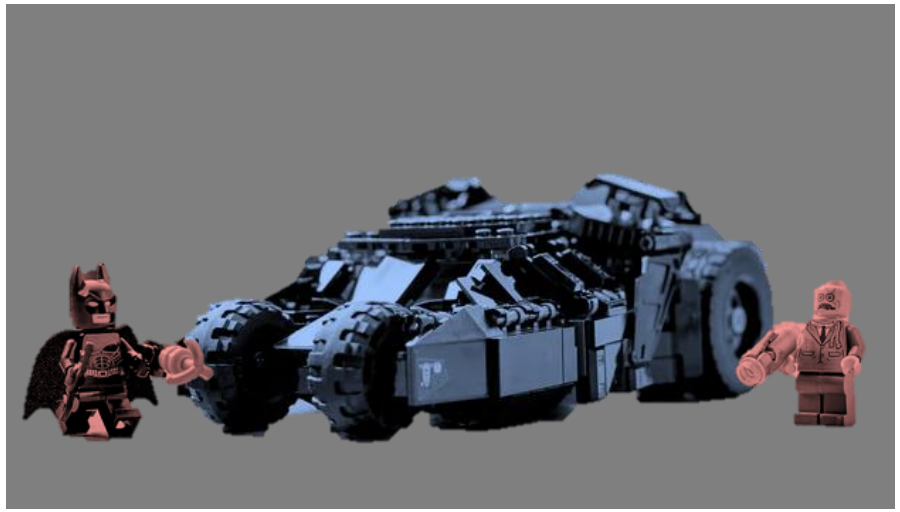
V. Conclusion

Ⅰ. Introduction

❖ Semantic segmentation 이란?

- 컴퓨터 비전(computer vision) 분야에서 다루는 문제 중 하나로 **이미지내의 모든 픽셀에 대해 클래스를 분류**하여 **의미 있는**(semantic) 단위로 **대상 객체를 분할**(segmentation)하는 문제

https://www.lego.com/ko-kr/product/lego-dc-batman-batmobile-tumbler-scarecrow-showdown-76239

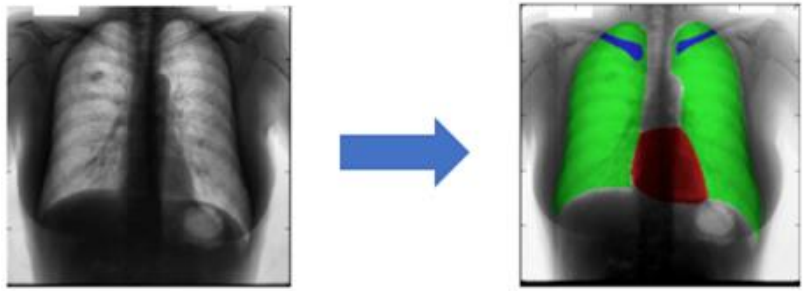


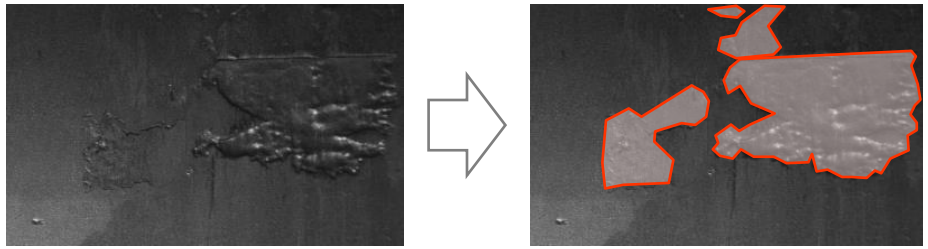
Ⅰ. Introduction

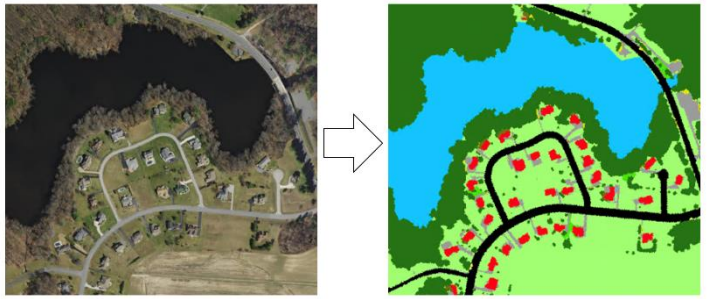
❖ Semantic segmentation의 활용

- 높은 수준의 이미지 인식을 가능하게 하여 **자율 주행, 의료 영상 분석, 표면 결함 검출, 위성 영상 분석 등** 다양한 산업에 활용

ICNet, Zhao et al. 2017: semantic segmentation demo video.







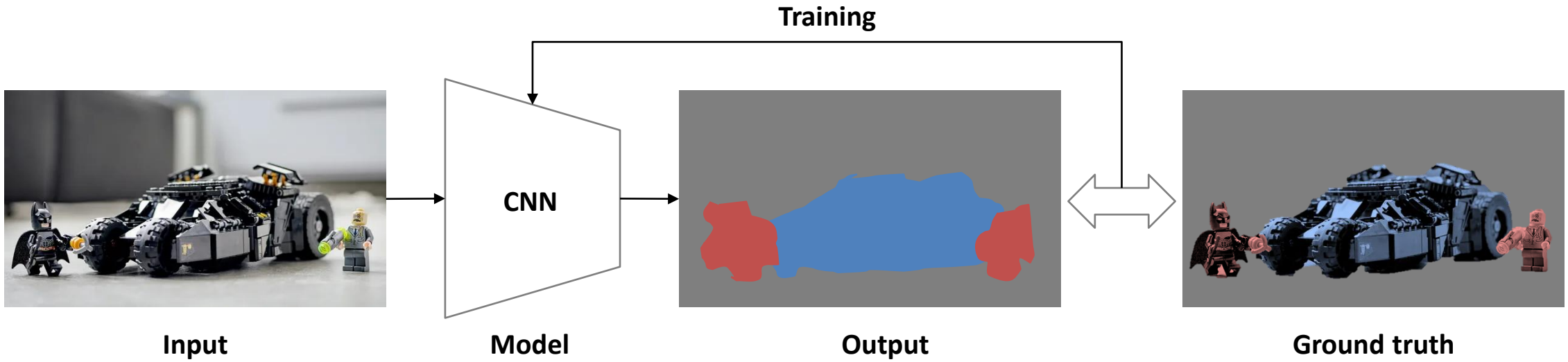
https://bskyvision.com/491 , https://thegradient.pub/semantic-segmentation/ , https://developers.arcgis.com/python/guide/how-unet-works/

Novikov, A. A., Lenis, D., Major, D., Hladůvka, J., Wimmer, M., & Bühler, K. (2018). Fully convolutional architectures for multiclass segmentation in chest radiographs. IEEE transactions on medical imaging, 37(8), 1865-1876.

Ⅰ. Introduction

❖ Semantic segmentation 모델 학습

- Semantic segmentation을 위한 딥러닝 모델을 학습시키기 위해서는 일반적으로 **입력 이미지와 그에 맞는 정답(label)이 필요**
- 딥러닝 모델로부터 얻은 **예측 결과와 정답을 비교**하여 모델 학습 과정을 반복
- 하지만, 딥러닝 모델은 점점 더 많은 데이터 수집을 필요로 하고 있어 **픽셀 단위의 정답 데이터를 얻기는 쉽지 않음**

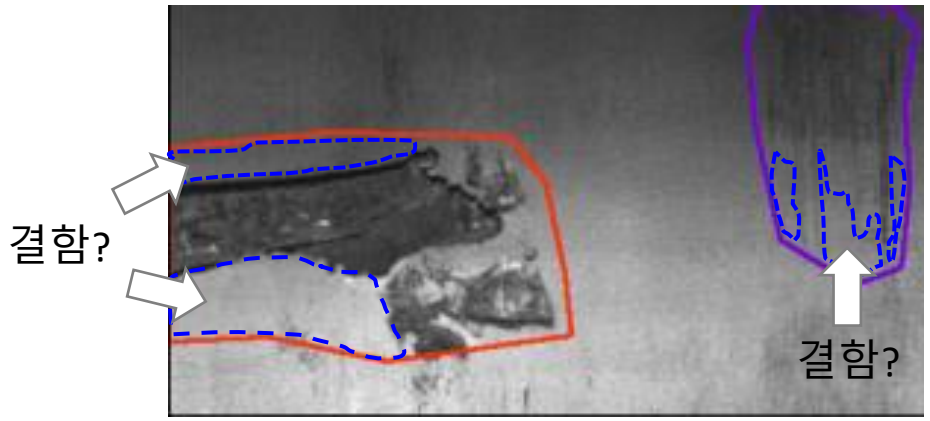


Ⅰ. Introduction

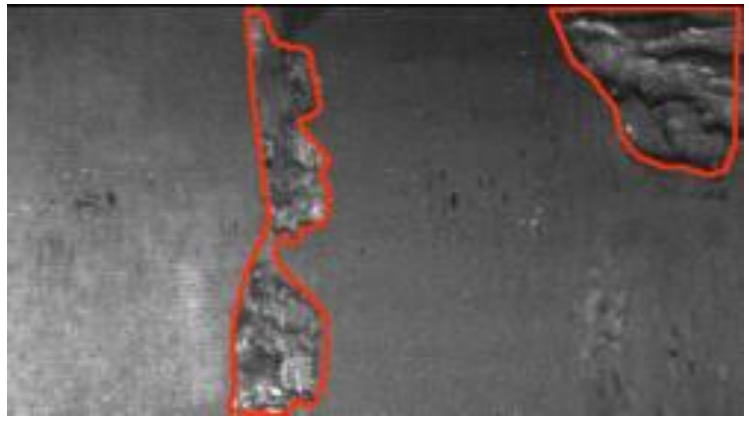
❖ Semantic segmentation을 위한 정답 데이터 구축

- Semantic segmentation은 이미지 수준의 정답이 아닌 픽셀 단위의 정답이 필요하기 때문에 데이터 구축이 더 어려움
- 특히, 데이터 구축 과정에서 잘못된 정답이 포함되기 쉬워 모델 학습에 오히려 악영향을 미칠 수 있음

강판 표면 결함 데이터 (Severstal 데이터셋)



<잘못된 영역이 포함된 정답 데이터>



<정확히 구축된 정답 데이터>

https://www.kaggle.com/c/severstal-steel-defect-detection/overview

Ⅰ. Introduction

❖ Approaches

- 정답 데이터 구축이 어려운 상황에서 접근 방법은 크게 **Weakly-supervised, Semi-supervised, Unsupervised**로 구분
- 본 세미나에서는 정답 데이터가 전혀 주어지지 않는 **Unsupervised semantic segmentation**에 대해 다룸

Weakly-supervised

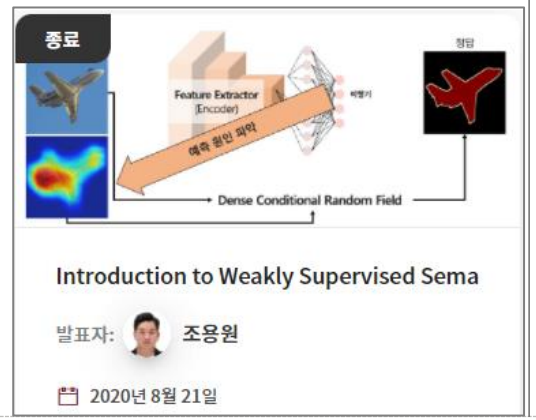
- Semantic segmentation에 대한 완전한 정답이 아닌 대략적인 마스크 또는 이미지 레이블 등 상대적으로 **적은 정보를 사용**하여 모델 학습

X

Y

… car … person …

0	1	0	0	1	0	0	0	0	0



"Introduction to Weakly Supervised Semantic Segmentation"
('20.8.21, 조용원 연구원)

Semi-supervised

- 전체 데이터셋 중 일부 소수의 데이터에만 정답이 주어지고 **다수의 정답이 없는 데이터를 함께 활용**하여 모델을 학습

X

Y

https://www.lego.com/ko-kr/themes/marvel

Unsupervised

- 모델 학습에 정답 데이터가 전혀 주어지지 않으며 주어진 **입력 이미지만으로** 모델을 학습

X

Y

I. Introduction

II. Unsupervised semantic segmentation

III. Mutual information maximization

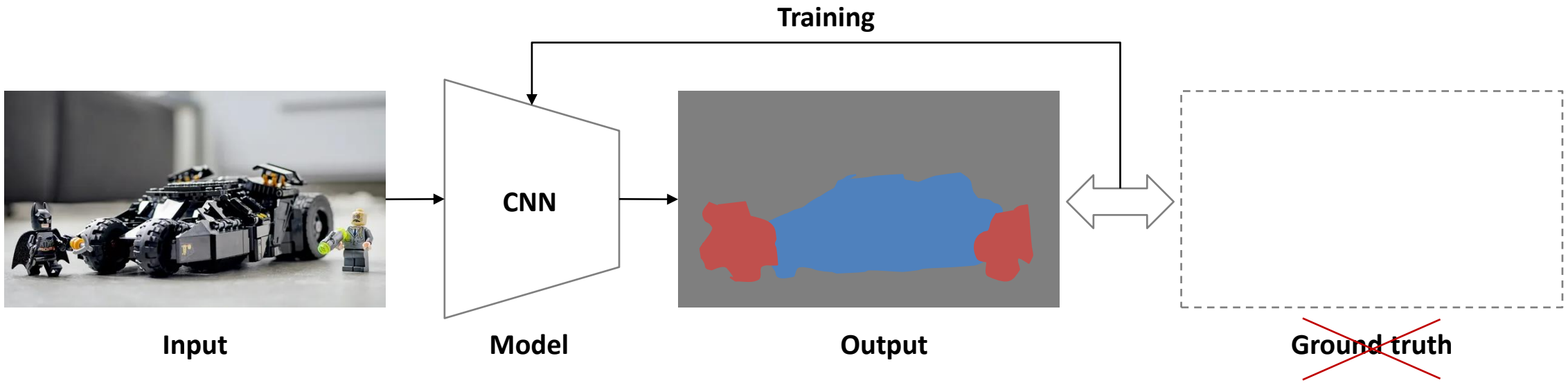
IV. Methods

V. Conclusion

Ⅱ. Unsupervised semantic segmentation

❖ Unsupervised learning

- 비지도 학습은 정답이 주어져 있는 않은 상황에서 **입력 데이터만으로 모델을 학습**시키는 방법론
- 입력 데이터만으로 모델을 학습시키기 위한 다양한 방법론이 연구되고 있음 (Clustering, Autoencoder, GAN, Self-supervised, …)



Ⅱ. Unsupervised semantic segmentation

❖ Unsupervised learning for semantic segmentation

- Semantic segmentation에서 비지도 학습은 정답 마스크(mask)가 주어지지 않은 상황에서 입력 이미지만으로 모델을 학습
- 단, 학습시 **클래스의 개수는 지정**되어 주어지며, 테스트 단계에서는 **예측 결과와 레이블을 매핑**하여 모델을 평가

Figure 1: From these unannotated images, we would like a recognition system to discover the concepts of *house*, *grass*, *trees* and *sky*, and segment each image accordingly without any supervision.

Ground truth

Prediction	tree	grass	sky	house	sum
1	0	4	1	0	5
2	0	0	0	3	3
3	0	0	4	0	4
4	1	1	0	0	2
sum	1	5	5	3	14

	tree	grass	sky	house	sum
1	1	1	0	0	2
2	0	4	1	0	5
3	0	0	4	0	4
4	0	0	0	3	3
sum	1	5	5	3	14

One-to-one mapping by solving the linear assignment problem using Hungarian method(Kuhn, 1955)

Cho, J. H., Mall, U., Bala, K., & Hariharan, B. (2021). PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16794-16804).

Ⅱ. Unsupervised semantic segmentation

❖ Unsupervised learning for semantic segmentation

* IIC : **I**nvariant **I**nformation **C**lustering(Ji *et al*., 2019)

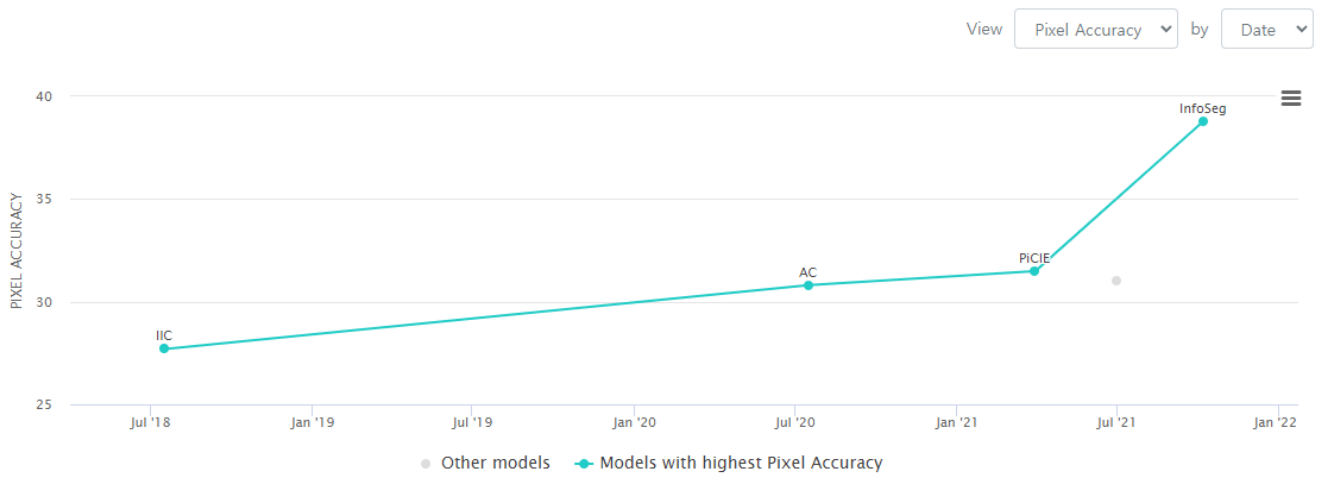
- 비지도 학습 기반의 Semantic segmentation은 **IIC**(Ji *et al*., 2019) **연구 이후 이를 baseline으로 한 연구들이 지속**되고 있음
- GAN 기반의 연구도 존재하지만 GAN 기반은 foreground와 background만을 구분할 수 있는 한계를 지님

GAN 기반 연구 : Labels4Free (Abdal *et al*., 2021)

Source Image | Background | Ground Truth | Our Mask | Difference | Our Composite

Unsupervised Semantic Segmentation on COCO-Stuff

Leaderboard Dataset



Rank	Model	Pixel Accuracy	Paper	Code	Result	Year
1	**InfoSeg**	38.8	InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization			2021
2	**PiCIE**	31.48	PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering			2021
3	**InMARS**	31.0	Unsupervised Image Segmentation by Mutual Information Maximization and Adversarial Regularization			2021
4	**AC**	30.8	Autoregressive Unsupervised Image Segmentation			2020
5	**IIC**	27.7	Invariant Information Clustering for Unsupervised Image Classification and Segmentation			2018

https://paperswithcode.com/sota/unsupervised-semantic-segmentation-on-coco-2

Abdal, R., Zhu, P., Mitra, N., & Wonka, P. (2021). Labels4Free: Unsupervised Segmentation using StyleGAN. arXiv preprint arXiv:2103.14968.

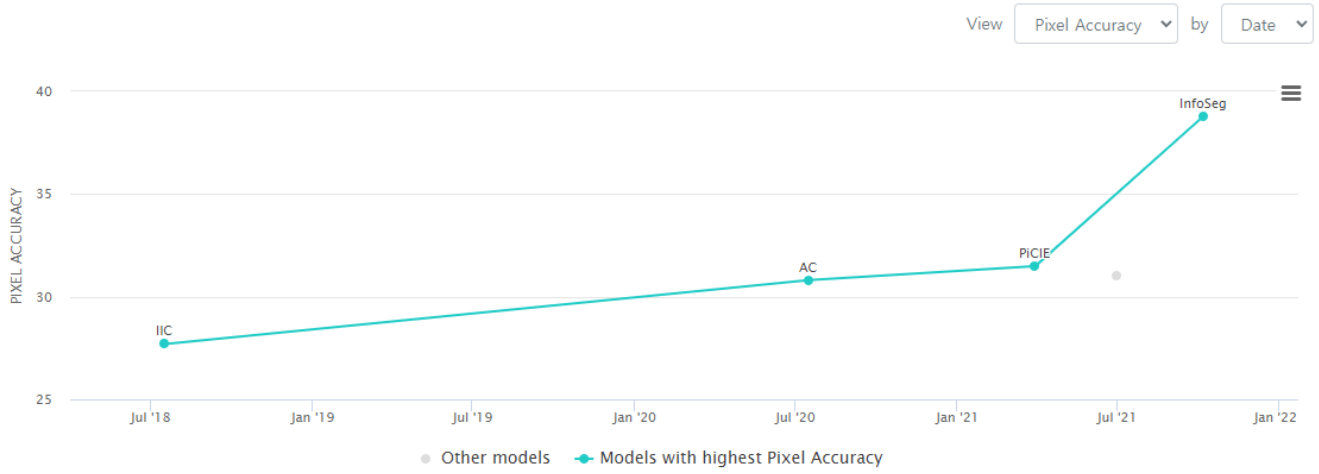
Ⅱ. Unsupervised semantic segmentation

❖ Unsupervised learning for semantic segmentation

- IIC(Ji *et al*., 2019) 이후의 연구들은 대부분 모델 학습의 objective로 **mutual information maximization**을 사용함
- 따라서 Unsupervised semantic segmentation의 최신 연구를 파악하기 위해서는 mutual information에 대한 이해가 필요

Unsupervised Semantic Segmentation on COCO-Stuff

Leaderboard Dataset



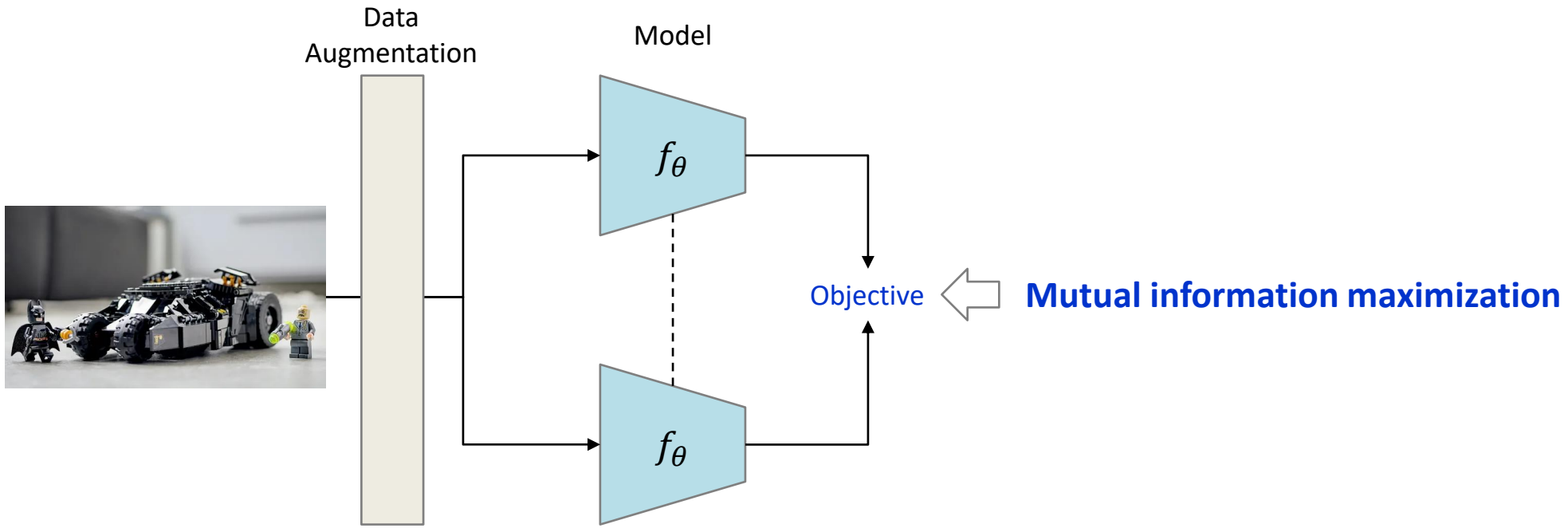
Rank	Model	Pixel Accuracy	Paper	Code	Result	Year
1	**InfoSeg**	38.8	InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization			2021
2	**PiCIE**	31.48	PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering			2021
3	**InMARS**	31.0	Unsupervised Image Segmentation by Mutual Information Maximization and Adversarial Regularization			2021
4	**AC**	30.8	Autoregressive Unsupervised Image Segmentation			2020
5	**IIC**	27.7	Invariant Information Clustering for Unsupervised Image Classification and Segmentation			2018

https://paperswithcode.com/sota/unsupervised-semantic-segmentation-on-coco-2

Ⅱ. Unsupervised semantic segmentation

❖ Unsupervised learning for semantic segmentation

- **입력 이미지와 증강된 이미지로 positive pair**를 구성하고 이를 모델에 통과시킨 후 이에 대한 **mutual information을 최대화**하는 목적식을 통해 모델을 학습



I. Introduction

II. Unsupervised semantic segmentation

III. Mutual information maximization

IV. Methods

V. Conclusion

Ⅲ. Mutual information maximization

❖ Mutual information (상호의존정보)

- **두 확률변수 사이의 상호의존성을 측정**한 것으로, 한 변수를 통해 얻어지는 다른 한 변수에 대한 정보량을 의미
- 두 확률변수의 독립여부를 측정할 수 있어 상관계수와 유사하게 **두 변수사이의 관계를 나타내는 척도**로 사용

정의

$$I(X;Y) \triangleq D_{KL}(P(x,y) \parallel P(x)P(y)) = \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}$$

두 확률분포 $P(x,y)$ 와 $P(x)P(y)$ 의 차이

차이 ↓ : $P(x,y) = P(x)P(y)$ ➔ *X*와 *Y*는 **독립관계** (상호의존성이 **낮다**)

차이 ↑ : $P(x,y) \neq P(x)P(y)$ ➔ *X*와 *Y*는 **의존관계** (상호의존성이 **높다**)

두 확률분포 P 와 Q 의 차이

$$D_{KL}(P \parallel Q) = \sum_{x\in X} P(x)\log\frac{P(x)}{Q(x)}$$

두 확률변수 *X*와 *Y*의 Mutual information을 높이는 방향으로 모델을 학습하면 두 변수의 상호의존성이 높아진다

https://en.wikipedia.org/wiki/Mutual_information

Ⅲ. Mutual information maximization

❖ Mutual information (상호의존정보)

- **두 확률변수 사이의 상호의존성을 측정**한 것으로, 한 변수를 통해 얻어지는 다른 한 변수에 대한 정보량을 의미
- 두 확률변수의 독립여부를 측정할 수 있어 상관계수와 유사하게 **두 변수사이의 관계를 나타내는 척도**로 사용

Entropy *H*

$$H(X) = -\sum_{x} P(x)\log(P(x))$$

정의

$$I(X;Y) \triangleq D_{KL}(P(x,y) \parallel P(x)P(y)) = \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}$$

$$= H(X) + H(Y) - H(X,Y)$$
$$= H(X) - H(X|Y)$$
$$= H(Y) - H(Y|X)$$
$$= I(Y;X)$$

두 확률분포 $P(x,y)$ 와 $P(x)P(y)$ 의 차이

차이 ↓ : $P(x,y) = P(x)P(y)$ ➔ *X*와 *Y*는 **독립관계** (상호의존성이 **낮다**)

차이 ↑ : $P(x,y) \neq P(x)P(y)$ ➔ *X*와 *Y*는 **의존관계** (상호의존성이 **높다**)

두 확률변수 *X*와 *Y*의 **Mutual information을 높이는 방향**으로 모델을 학습하면 두 변수의 상호의존성이 높아진다

: 변수 *X*의 불확실성($H(X)$)

: 변수 *Y*가 주어졌을 때 변수 *X*의 불확실성($H(X|Y)$)

Mutual information은 한 변수를 통해 다른 한 변수의 불확실성이 감소되는 정보량을 나타낸다

https://en.wikipedia.org/wiki/Mutual_information

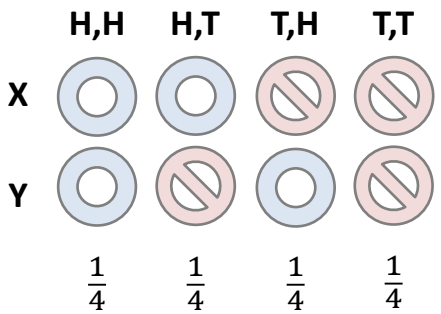
Ⅲ. Mutual information maximization

$$I(X;Y) \triangleq D_{KL}(P(x,y) \parallel P(x)P(y)) = \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \begin{array}{l} = H(X) + H(Y) - H(X,Y) \\ = H(X) - H(X|Y) \\ = H(Y) - H(Y|X) \end{array}$$

❖ Mutual information - Example

- **두 확률변수 사이의 상호의존성을 측정**한 것으로, 한 변수를 통해 얻어지는 다른 한 변수에 대한 정보량을 의미
- 두 확률변수의 독립여부를 측정할 수 있어 상관계수와 유사하게 **두 변수사이의 관계를 나타내는 척도**로 사용

Case 1. X,Y가 독립관계



X \ Y	0	1	
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

$$\begin{aligned} I(X;Y) &= \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \\ &= H(X) - H(X|Y) \\ &= -2 \times \frac{1}{2}\log\left(\frac{1}{2}\right) - \left(\frac{1}{2}H(X|Y=0) + \frac{1}{2}H(X|Y=1)\right) \\ &= 1 - 1 = \mathbf{0} \end{aligned}$$

[해석]

1. Y변수가 주어졌을 때 X변수의 불확실성이 감소되는 양은 0이다.
2. X와 Y의 상호정보량이 0이므로 X와 Y는 독립관계이다.

https://www.youtube.com/watch?v=U9h1xkNELvY&t=327s , https://bskyvision.com/774 , http://www.cs.tau.ac.il/~iftachh/Courses/Info/Fall14/Printouts/Lesson2_h.pdf

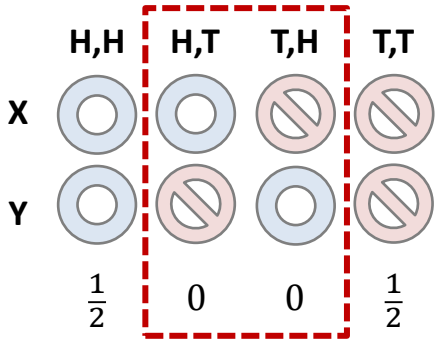
Ⅲ. Mutual information maximization

$$I(X;Y) \triangleq D_{KL}(P(x,y) \parallel P(x)P(y)) = \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \begin{array}{l} = H(X) + H(Y) - H(X,Y) \\ = H(X) - H(X|Y) \\ = H(Y) - H(Y|X) \end{array}$$

❖ Mutual information - Example

- **두 확률변수 사이의 상호의존성을 측정**한 것으로, 한 변수를 통해 얻어지는 다른 한 변수에 대한 정보량을 의미
- 두 확률변수의 독립여부를 측정할 수 있어 상관계수와 유사하게 **두 변수사이의 관계를 나타내는 척도**로 사용

Case 2. X,Y가 의존관계(1)



X \ Y	0	1	
0	1/2	0	1/2
1	0	1/2	1/2
	1/2	1/2	

$$\begin{aligned} I(X;Y) &= \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \\ &= H(X) - H(X|Y) \\ &= -2\times\frac{1}{2}\log\left(\frac{1}{2}\right) - \left(\frac{1}{2}H(X|Y=0) + \frac{1}{2}H(X|Y=1)\right) \\ &= 1 - 0 = \mathbf{1} \end{aligned}$$

[해석]

1. Y변수가 주어졌을 때 X변수의 불확실성은 모두 감소된다.
2. X와 Y의 상호정보량이 1이므로 X와 Y는 완전한 의존관계이다.

https://www.youtube.com/watch?v=U9h1xkNELvY&t=327s , https://bskyvision.com/774 , http://www.cs.tau.ac.il/~iftachh/Courses/Info/Fall14/Printouts/Lesson2_h.pdf

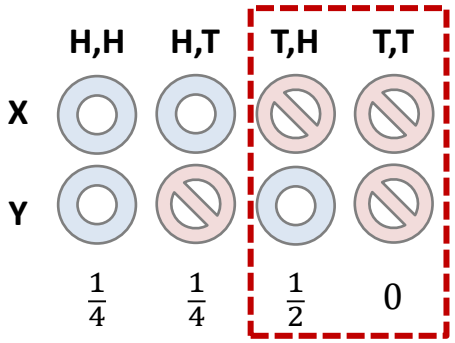
Ⅲ. Mutual information maximization

$$I(X;Y) \triangleq D_{KL}(P(x,y) \parallel P(x)P(y)) = \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \begin{array}{l} = H(X) + H(Y) - H(X,Y) \\ = H(X) - H(X|Y) \\ = H(Y) - H(Y|X) \end{array}$$

❖ Mutual information - Example

- **두 확률변수 사이의 상호의존성을 측정**한 것으로, 한 변수를 통해 얻어지는 다른 한 변수에 대한 정보량을 의미
- 두 확률변수의 독립여부를 측정할 수 있어 상관계수와 유사하게 **두 변수사이의 관계를 나타내는 척도**로 사용

Case 3. X,Y가 의존관계(2)



X \ Y	0	1	
0	1/4	1/4	1/2
1	1/2	0	1/2
	3/4	1/4	

$$\begin{aligned} I(X;Y) &= \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \\ &= H(X) - H(X|Y) \\ &= -2 \times \frac{1}{2}\log\left(\frac{1}{2}\right) - \left(\frac{3}{4}H(X|Y=0) + \frac{1}{4}H(X|Y=1)\right) \\ &= 1 - 0.6887 = \mathbf{0.3113} \end{aligned}$$

[해석]

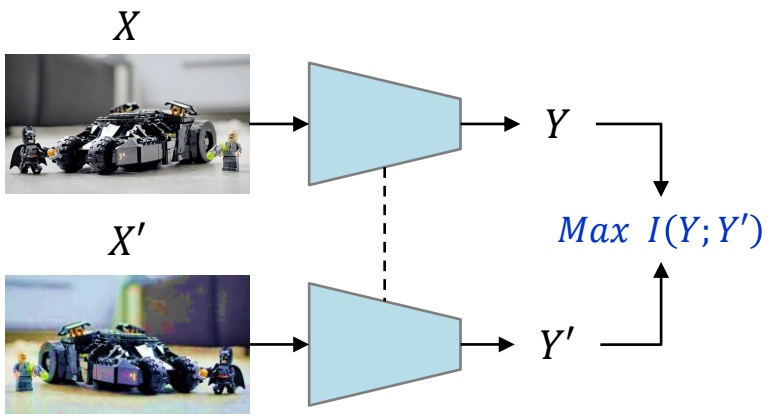
1. Y변수가 주어졌을 때 X변수의 불확실성은 0.3113 감소된다.
2. X와 Y는 0.3113만큼의 상호정보량을 가지는 의존관계에 있다.

https://www.youtube.com/watch?v=U9h1xkNELvY&t=327s , https://bskyvision.com/774 , http://www.cs.tau.ac.il/~iftachh/Courses/Info/Fall14/Printouts/Lesson2_h.pdf

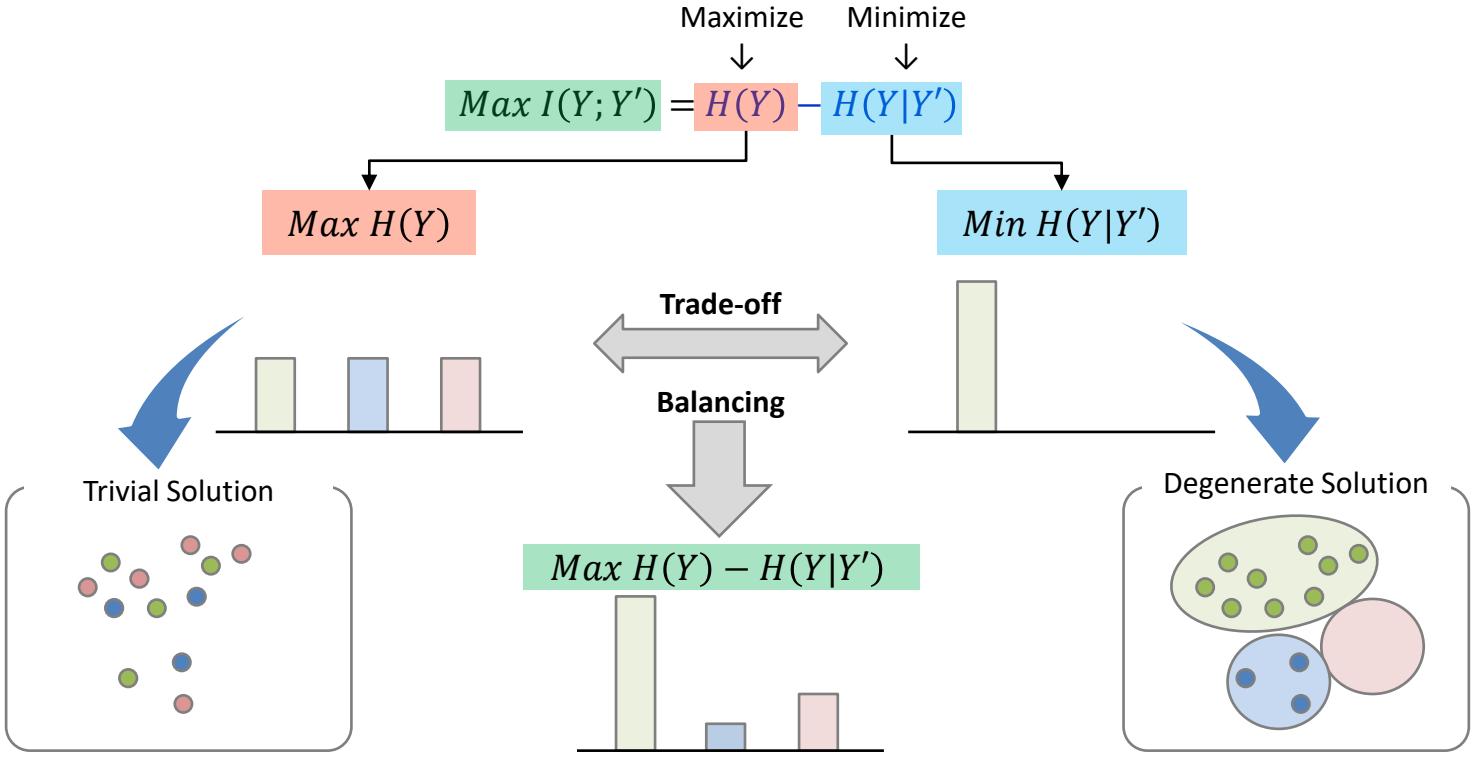
Ⅲ. Mutual information maximization

❖ Why mutual information?

$$Max\ I(Y;Y') = D_{KL}(P(y,y') \parallel P(y)P(y'))$$



- MI 최대화를 통해 두 이미지 사이의 **공통된 특징이 잘 추출되는 방향**으로 모델을 학습



- $H(Y)$**는 최대화,** $H(Y|Y')$**은 최소화하도록 균형**을 이루면서 clustering에서 발생할 수 있는 degenerate solution과 같은 문제를 회피하는 모델 학습

Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).
Alguwaizani, A. (2012). Degeneracy on K-means clustering. Electronic Notes in Discrete Mathematics, 39, 13-20.

I. Introduction

II. Unsupervised semantic segmentation

III. Mutual information maximization

IV. Methods

V. Conclusion

Rank	Model	Pixel Accuracy	Paper	Code	Result	Year
1	**InfoSeg**	38.8	InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization			2021
2	**PiCIE**	31.48	PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering			2021
3	**InMARS**	31.0	Unsupervised Image Segmentation by Mutual Information Maximization and Adversarial Regularization			2021
4	**AC**	30.8	Autoregressive Unsupervised Image Segmentation			2020
5	**IIC**	27.7	Invariant Information Clustering for Unsupervised Image Classification and Segmentation			2018

https://paperswithcode.com/sota/unsupervised-semantic-segmentation-on-coco-2

IV. Methods – 1) IIC (Invariant Information Clustering)

❖ Invariant Information Clustering for Unsupervised Image Classification and Segmentation

- Ji *et al.* (University of Oxford), 2019 International Conference on Computer Vision (ICCV)
- 364회 인용 ('22.3.18 기준)

Invariant Information Clustering for
Unsupervised Image Classification and Segmentation

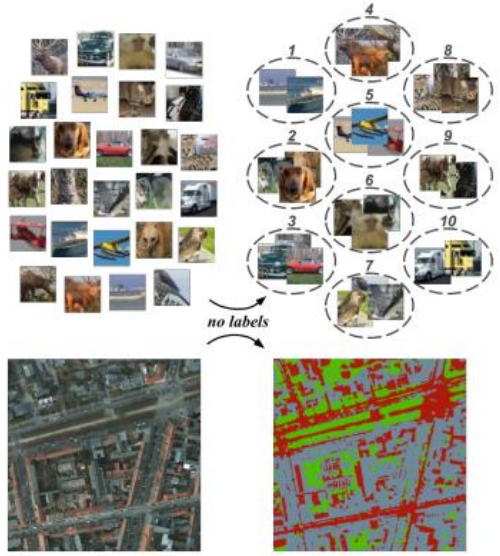
Xu Ji
University of Oxford
xuji@robots.ox.ac.uk

João F. Henriques
University of Oxford
joao@robots.ox.ac.uk

Andrea Vedaldi
University of Oxford
vedaldi@robots.ox.ac.uk

Abstract

We present a novel clustering objective that learns a neural network classifier from scratch, given only unlabelled data samples. The model discovers clusters that accurately match semantic classes, achieving state-of-the-art results in eight unsupervised clustering benchmarks spanning image classification and segmentation. These include STL10, an unsupervised variant of ImageNet, and CIFAR10, where we significantly beat the accuracy of our closest competitors by 6.6 and 9.5 absolute percentage points respectively. The method is not specialised to computer vision and operates on any paired dataset samples; in our experiments we use random transforms to obtain a pair from each image. The trained network directly outputs semantic labels, rather than high dimensional representations that need external processing to be usable for semantic clustering. The objective is simply to maximise mutual information between the class assignments of each pair. It is easy to implement and rigorously grounded in information theory, meaning we effortlessly avoid degenerate solutions that other clustering



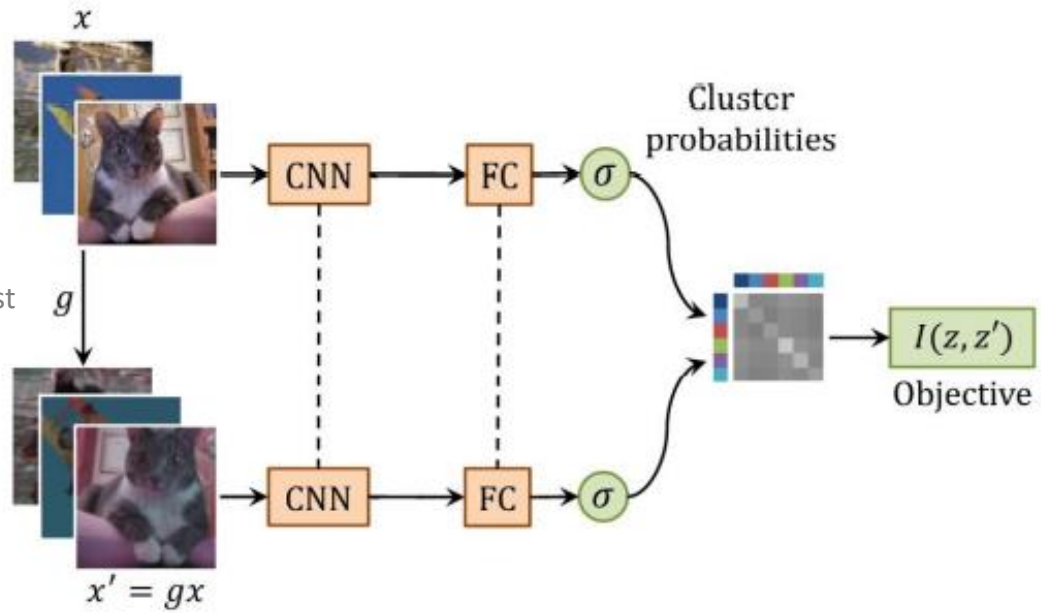
Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).

IV. Methods – 1) IIC (Invariant Information Clustering)

❖ IIC approach

- **입력 이미지 → CNN → FC → Output**(cluster probability) 의 단순한 모델 구조
- 입력 이미지에 data augmentation을 적용하여 **positive pair를 생성**하고 동일한 모델에 통과시켜 예측 결과를 얻음(z, z')
- 두 예측결과인 z와 z' 사이의 **mutual information이 최대화**되도록 모델 학습

g : scaling, skewing, rotation, flipping, contrast and color saturation



Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).

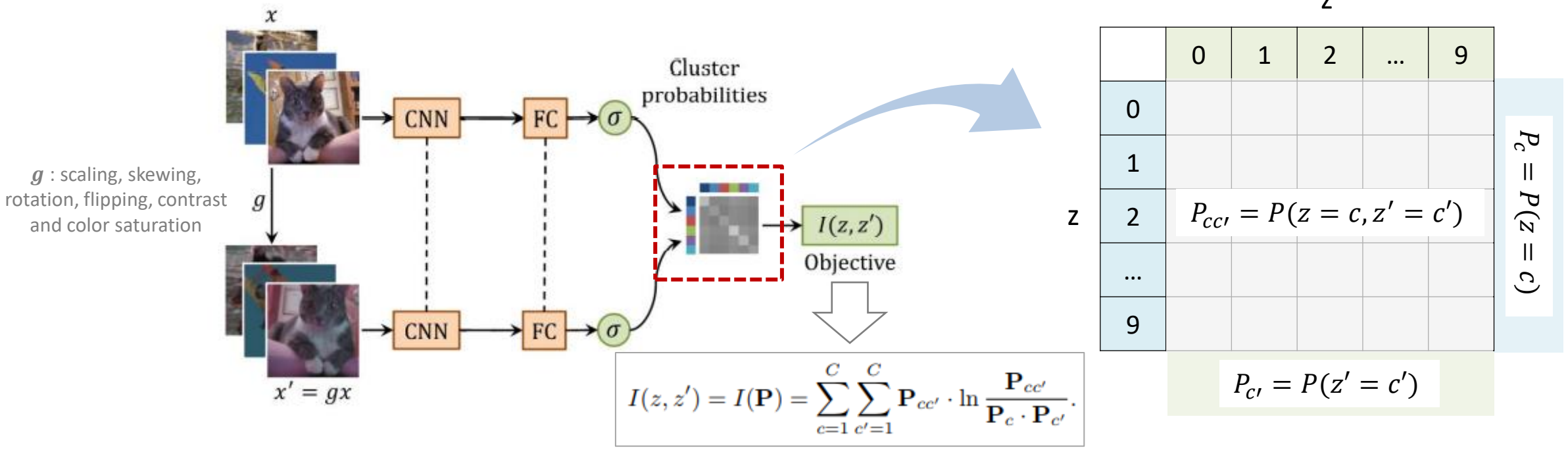
IV. Methods – 1) IIC (Invariant Information Clustering)

❖ Mutual information

- **Mutual information을 최대화**하여 두 이미지간의 공통 특징을 최대한 추출하도록 학습하고 이 정보를 기반으로 clustering

  ➔ **Invariant information clustering**

- Mutual information($I(z, z') = H(z) - H(z|z')$) 최대화는 **$H(z)$를 최대화하면서 동시에 $H(z|z')$을 최소화**하는 것으로 clustering에서의 degenerate solution과 같은 문제를 해결

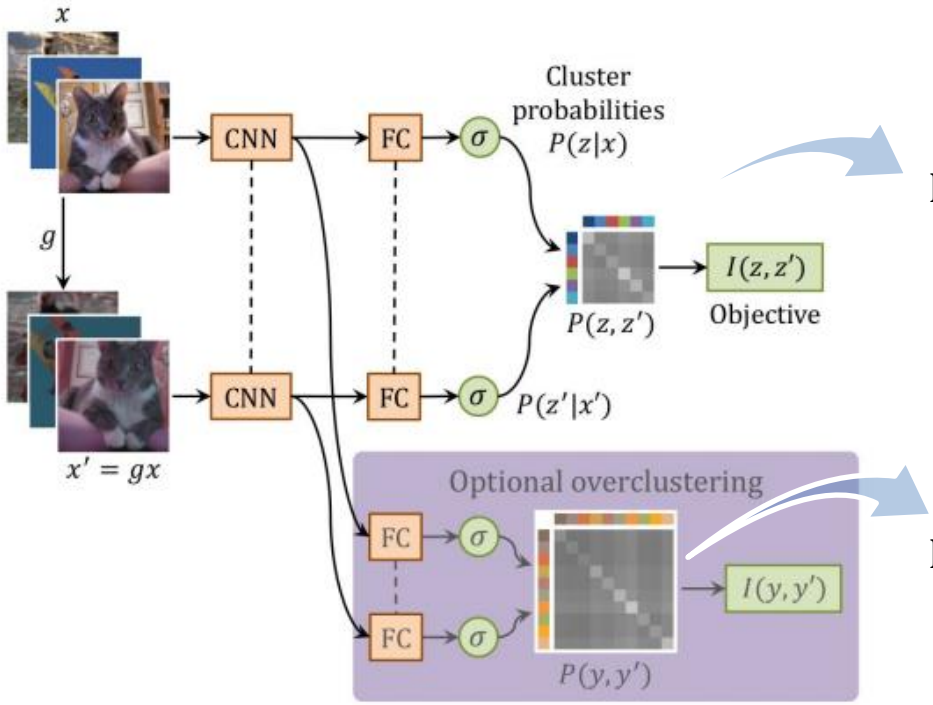


Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).

IV. Methods – 1) IIC (Invariant Information Clustering)

❖ Overclustering

- 목표 클래스 수보다 더 많은 수의 cluster를 가진 **overclustering** head를 추가하여 학습
- 목표 클래스와 관련 없는 이미지까지 최대한 활용하기 위해 제안된 방법으로, 이미지의 더 좋은 특징을 추출할 수 있도록 학습에 활용함 (e.g., STL10 dataset : 목표 클래스 10개에 대한 데이터 13K개 + 목표 클래스 이외 데이터 100K개)



Number of Cluster $k = k_{gt}$

Number of Cluster $k > k_{gt}$

$k_{gt} = 10, k_{overcluster} = 70$

+15.8%p

	STL10
No auxiliary overclustering	43.8
Single sub-head ($h = 1$)	57.6
No sample repeats ($r = 1$)	47.0
Unlabelled data segment ignored	49.9
Full setting	**59.6**

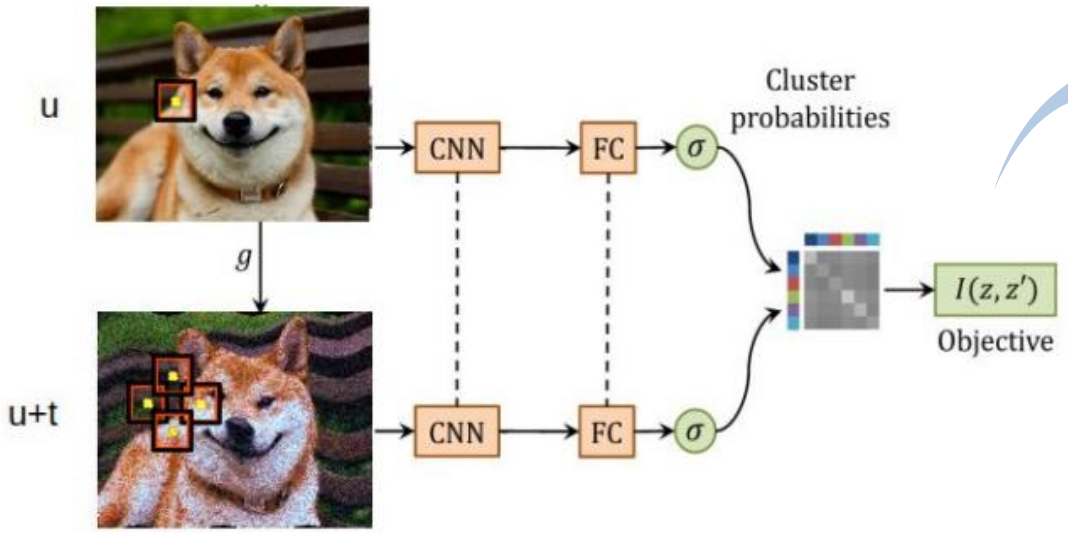
Table 2: **Ablations of IIC (unsupervised setting).** Each row shows a single change from the full setting. The full setting has auxiliary overclustering, 5 initialisation heads, 5 sample repeats, and uses the unlabelled data subset of STL10.

Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).

IV. Methods – 1) IIC (Invariant Information Clustering)

❖ Unsupervised semantic segmentation

- 이미지 단위의 분류와는 달리, 픽셀 단위의 분류를 위해 **이미지의 부분적인 패치 정보**를 사용
- 픽셀 단위로 **해당 픽셀을 중심으로 한 패치**와 증강된 이미지에서 ***t*만큼 이동한 패치**와의 mutual information을 최대화
- 즉, 근접 패치와의 **local spatial invariance**를 가정하고 한 픽셀에 대해 **주변과의 공통 정보를 최대화**하면서 모델학습



$$\max_{\Phi} \frac{1}{|T|} \sum_{t \in T} I(\mathbf{P}_t),$$

$$\mathbf{P}_t = \frac{1}{n|G||\Omega|} \sum_{i=1}^{n} \sum_{g \in G} \overbrace{\sum_{u \in \Omega} \Phi_u(\mathbf{x}_i) \cdot [g^{-1}\Phi(g\mathbf{x}_i)]_{u+t}^{\top}}^{\text{Convolution}}.$$

픽셀단위 예측결과 → z

z′

	0	1	2	…	9
0					
1					
2			$P_{cc'} = P(z = c, z' = c')$		
…					
9					

$P_c = P(z = c)$

$P_{c'} = P(z' = c')$

https://github.com/vandedok/IIC_tutorial

Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).

IV. Methods – 1) IIC (Invariant Information Clustering)

❖ Experiments (Unsupervised semantic segmentation)

- **Dataset**

	COCO-Stuff-15		COCO-Stuff-3		Potsdam-6		Potsdam-3	
	Train	Test	Train	Test	Train	Test	Train	Test
IIC	51804	51804	36660	36660	8550	5400	8550	5400
Semi-supervised	49629	2175	35228	1432	7695	855	7695	855

Table 6: Datasets for segmentation.

- **Hyperparameters**

Overclustering ↓

	b	n	h	r	k_{in}	k_{gt}	k	crop size(s)	input size
COCO-Stuff-3	C	120	1	1	5	3	15	128	128
COCO-Stuff	C	60	1	1	5	15	45	128	128
Potsdam-3	C	75	1	1	4	3	24	200	200
Potsdam	C	60	1	1	4	6	36	200	200

- **Network**

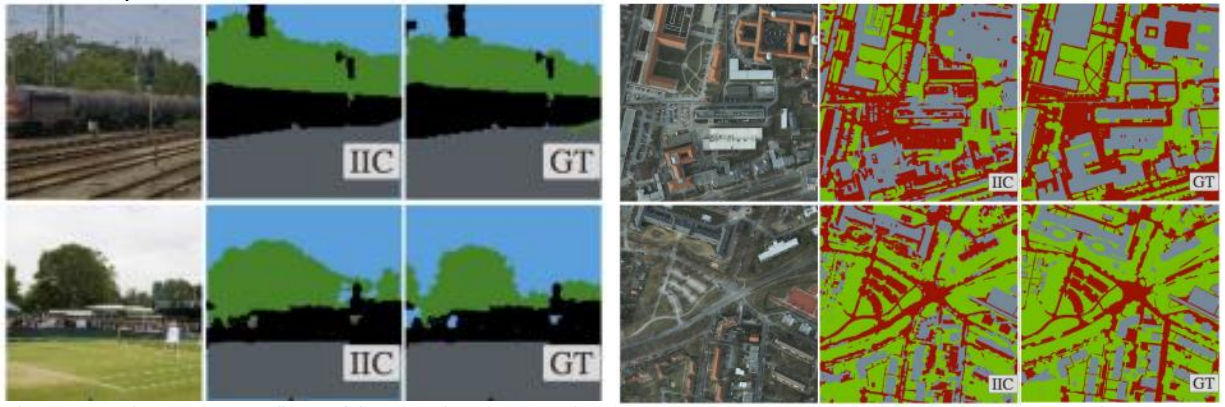
C
1× Conv@64
1× Conv@128
1× MaxPool
2× Conv@256
2× Conv@512

- **Result** (Pixel accuracy)

	COCO-Stuff-3	COCO-Stuff	Potsdam-3	Potsdam
Random CNN	37.3	19.4	38.2	28.3
K-means [44]†	52.2	14.1	45.7	35.3
SIFT [39]‡	38.1	20.2	38.2	28.5
Doersch 2015 [17]‡	47.5	23.1	49.6	37.2
Isola 2016 [30]‡	54.0	24.3	63.9	44.9
DeepCluster 2018 [7]† ‡	41.6	19.9	41.7	29.2
IIC	**72.3**	**27.7**	**65.1**	**45.4**

Table 4: **Unsupervised segmentation.** IIC experiments use a single sub-head. Legend: †Method based on k-means. ‡Method that does not directly learn a clustering function and requires further application of k-means to be used for image clustering.

* non-stuff pixels in black



Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9865-9874).

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Autoregressive Unsupervised Image Segmentation

- Ouali *et al.*, 2020 European Conference on Computer Vision (ECCV)
- 17회 인용 ('22.3.18 기준)

Autoregressive Unsupervised Image Segmentation

Yassine Ouali, Céline Hudelot and Myriam Tami

Université Paris-Saclay, CentraleSupélec, MICS, 91190, Gif-sur-Yvette, France
{yassine.ouali,celine.hudelot,myriam.tami}@centralesupelec.fr

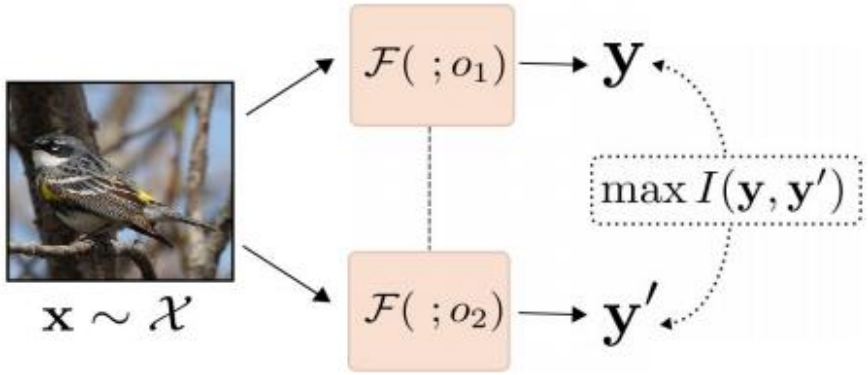
Abstract. In this work, we propose a new unsupervised image segmentation approach based on mutual information maximization between different constructed views of the inputs. Taking inspiration from autoregressive generative models that predict the current pixel from *past* pixels in a raster-scan ordering created with masked convolutions, we propose to use different *orderings* over the inputs using various forms of masked convolutions to construct different *views* of the data. For a given input, the model produces a pair of predictions with two valid orderings, and is then trained to maximize the mutual information between the two outputs. These outputs can either be low-dimensional features for representation learning or output clusters corresponding to semantic labels for clustering. While masked convolutions are used during training, in inference, no masking is applied and we fall back to the standard convolution where the model has access to the full input. The proposed method outperforms current state-of-the-art on unsupervised image segmentation. It is simple and easy to implement, and can be extended to other visual tasks and integrated seamlessly into existing unsupervised learning methods requiring different views of the data.

Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

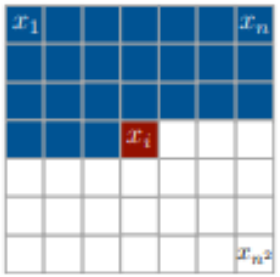
IV. Methods – 2) AC (Autoregressive Clustering)

❖ Autoregressive clustering

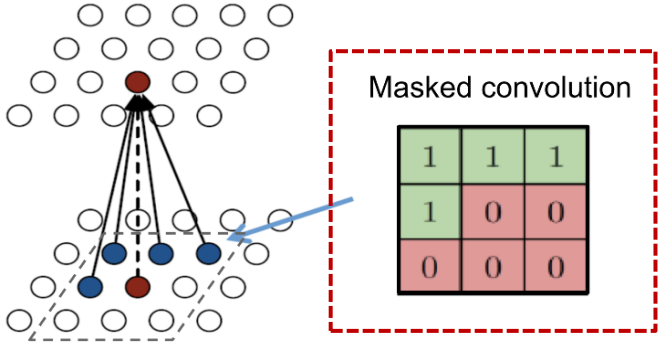
- 이미지 생성 방법론 중 autoregressive 모델인 PixelCNN의 **masked convolution**을 개선하여 segmentation 문제에 활용
- Masked convolution을 통해 입력 이미지에 대한 두 개의 view를 생성하고, 예측된 결과에 대해 **mutual information을 최대화**



Pixel Recurrent Neural Networks
(Van Oord *et al.*, 2016)



Context



PixelCNN

$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i|x_1, ..., x_{i-1})$$

https://github.com/anordertoreclaim/PixelCNN

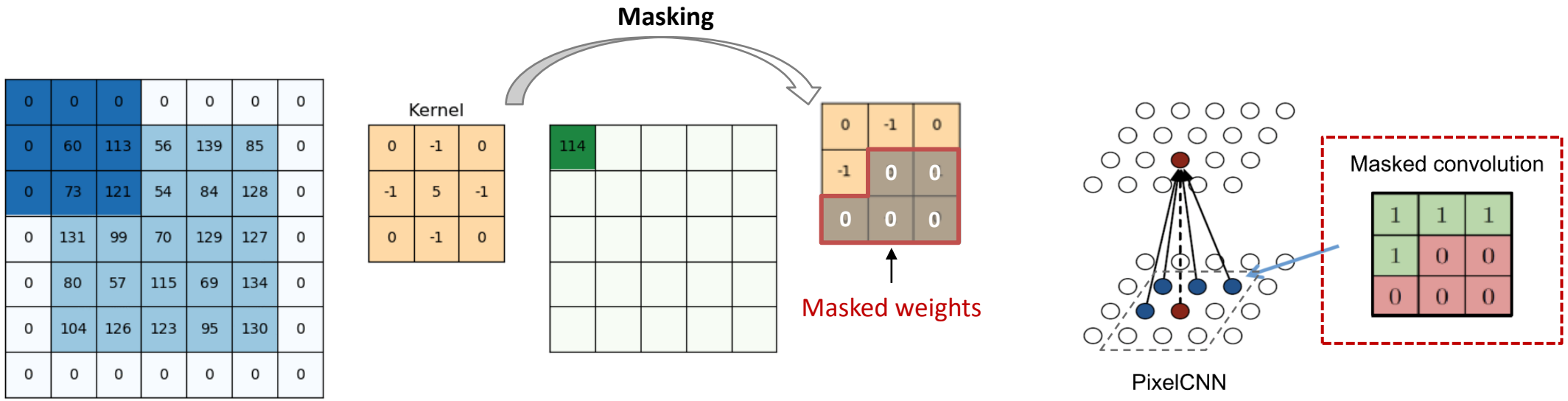
https://yassouali.github.io/autoreg_seg/

Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Masked convolution

- 일반적인 CNN에서 사용하는 **convolution filter의 weight에 masking(0)을 적용**하여 convolution 연산시 제외하고, masked weights는 모델 학습시 update를 하지 않음



https://oi.readthedocs.io/en/latest/computer_vision/cnn/intro.html

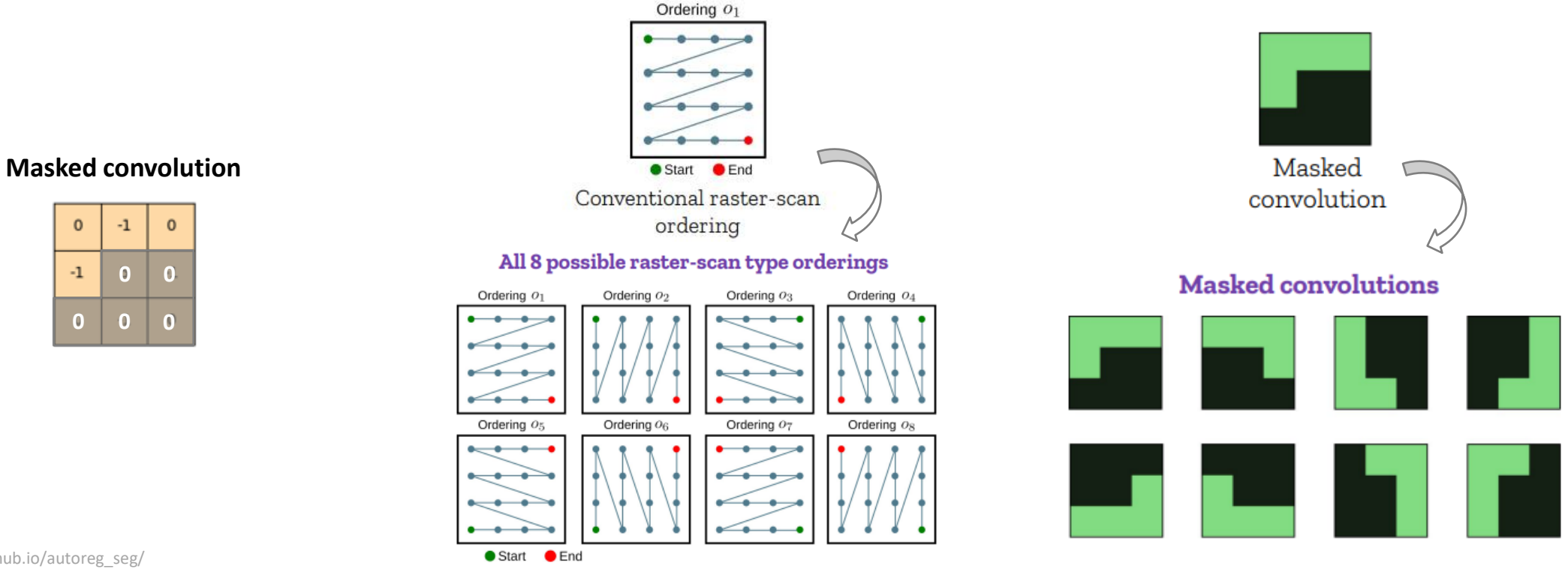
https://yassouali.github.io/autoreg_seg/

Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Masked convolution

- PixelCNN(Van Oord *et al.*, 2016)에서 제안된 기존의 masked convolution은 이미지의 왼쪽 위에서 시작하여 오른쪽 아래로 향하는 방향만을 가정하고 한가지 형태만을 가진 masked convolution이 제안됨
- AC에서는 **masked convolution의 방향과 형태를 다양하게 변형**하고 적용하여 원본 이미지에 대한 두 개의 view를 생성



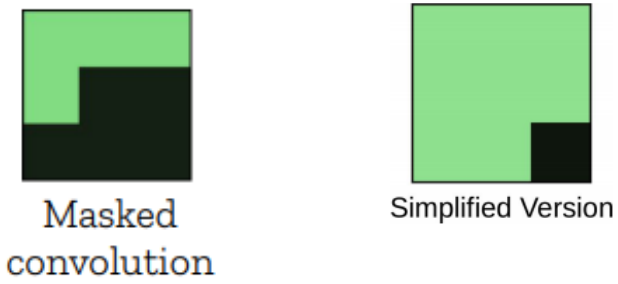
https://yassouali.github.io/autoreg_seg/

Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Masked convolution

- 하지만, masked weights가 너무 많으면 receptive field가 줄어들고 학습되는 weight의 수가 줄어들어 학습효과가 떨어지기 때문에 **단순화된 masked convolution**을 제안
- 또한, masked convolution의 특성상 한 픽셀의 receptive field에서 blind spot이 생기기 때문에 이를 해결하기 위해 **attention mechanism을 적용**하고, 이를 통해 최종적으로 **zigzag 타입의 masked convolution** 형태도 추가



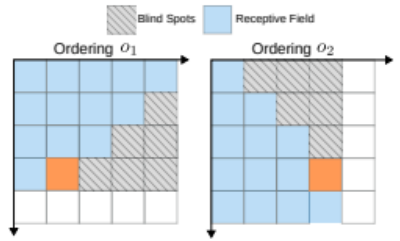


Fig. 6. Blind Spots. Blind spots in the receptive field of pixel ■ as a result of using a masked convolution for a given ordering o_i.

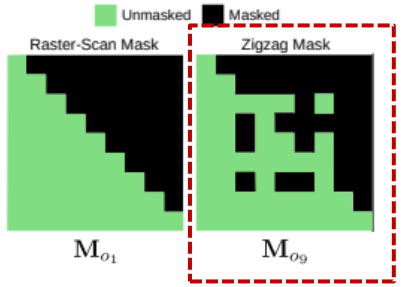


Fig. 5. Attention Masks. Examples of the different attention masks $\mathbf{M}_{o_i}$ of shape $HW \times HW$ applied for a given ordering o_i. With $HW = 9$.

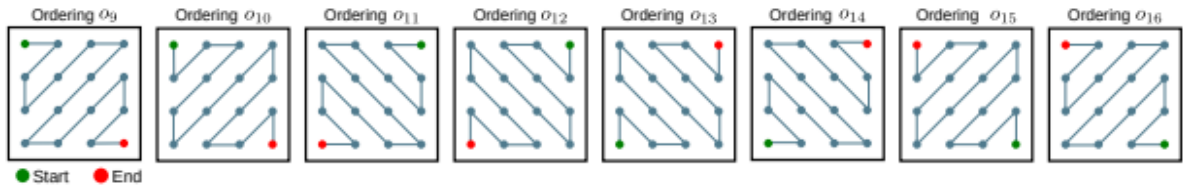


Fig. 4. Zigzag type orderings.

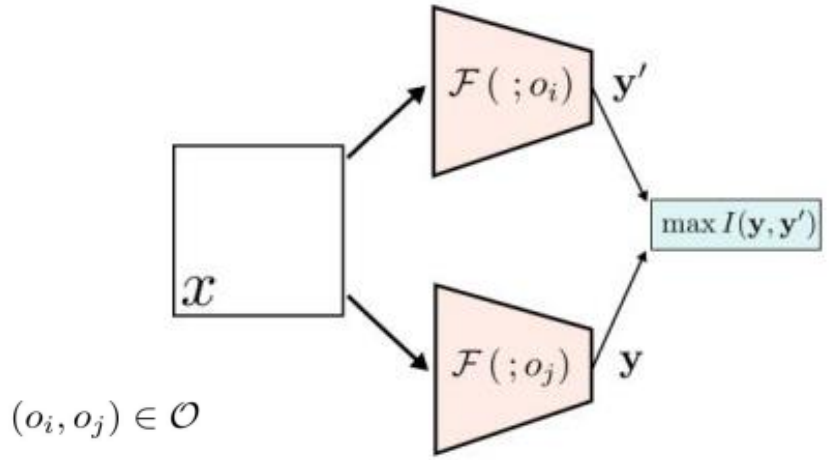
https://yassouali.github.io/autoreg_seg/

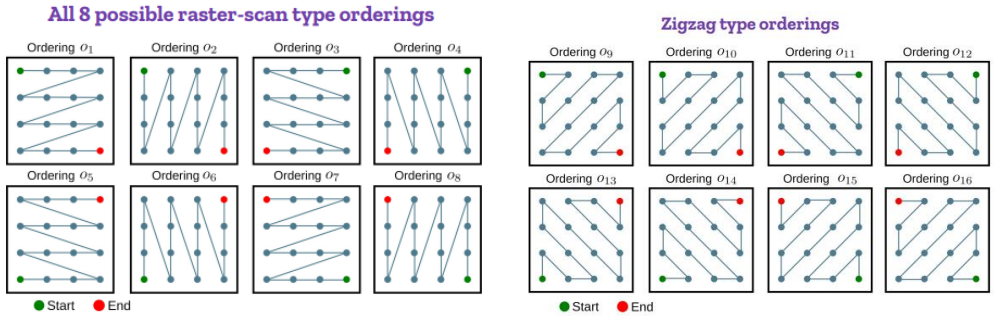
Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Training with masked convolution

- 다양한 masked convolution 형태 중 **두 개를 샘플링**하고 적용하여 원본 이미지에 대한 두 개의 view를 생성하고 **unmasked weight에 대해서만 update**를 실시





Training procedure

1- Sample two valid orderings $(o_i, o_j) \in \mathcal{O}$

2- Set the correct masking and padding for o_i

- Compute the output corresponding to the first view

3- Set the correct masking and padding for o_j

- Compute the output corresponding to the second view

4- Compute the loss

5- Backpropagate and only update the unmasked weights, masked weights remain unchanged

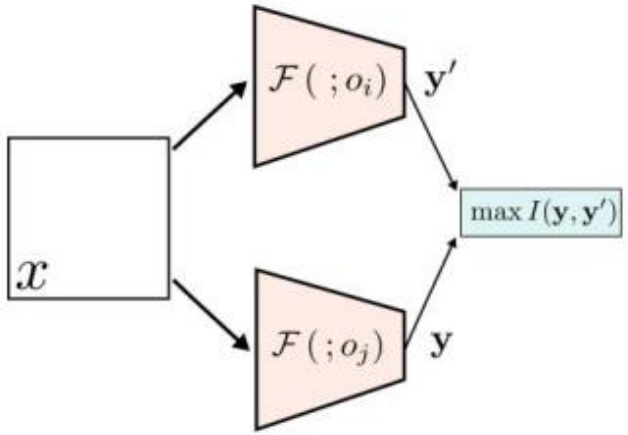
https://yassouali.github.io/autoreg_seg/

Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Mutual information maximization

- 두 개의 view로부터 나온 output 간의 **mutual information**을 objective function으로 사용

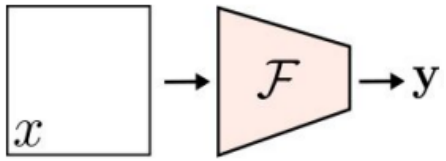


Maximizing the extact value
for clustering

$$\mathcal{L}_{\mathrm{AC}} = \mathbb{E}_{\mathbf{x}\sim\mathcal{X}}\left[\mathbb{E}_{p(\mathbf{y},\mathbf{y}')}\log\frac{p(\mathbf{y},\mathbf{y}')}{p(\mathbf{y})p(\mathbf{y}')}\right]$$

❖ Inference

- Inference 단계에서는 **masking 없이** 학습된 normal convolution을 적용하여 픽셀단위의 클래스 예측



Clustering
$\mathbf{y}$: output probabilities → argmax

https://yassouali.github.io/autoreg_seg/

Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 2) AC (Autoregressive Clustering)

❖ Experiments

- 실험 결과, 동일한 데이터셋에 대해 IIC 방법론 대비 좀 더 나은 성능의 결과를 보임
- 입력 이미지의 data augmentation 적용시 효과와 masked convolution의 종류가 다양할수록 성능이 향상됨을 보임

	COCO-Stuff-3	COCO-Stuff	Potsdam-3	Potsdam
Random CNN	37.3	19.4	38.2	28.3
K-means [42]	52.2	14.1	45.7	35.3
SIFT [35]	38.1	20.2	38.2	28.5
Doersch 2015 [11]	47.5	23.1	49.6	37.2
Isola 2016 [27]	54.0	24.3	63.9	44.9
DeepCluster 2018 [6]	41.6	19.9	41.7	29.2
IIC 2019 [28]	72.3	27.7	65.1	45.4
AC	**72.9**	**30.8**	**66.5**	**49.3**

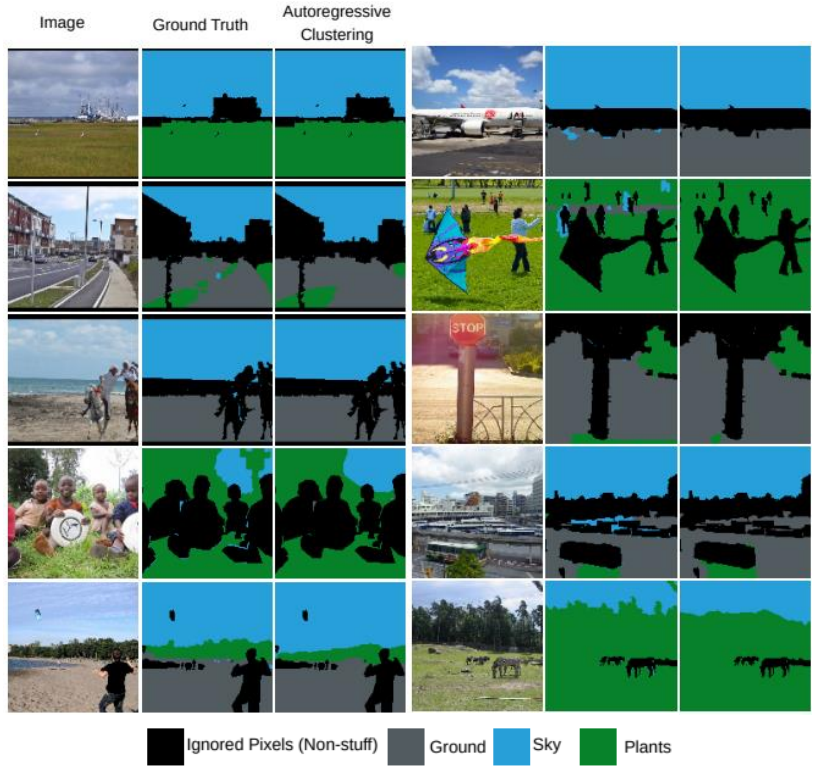
Table 3. Unsupervised image segmentation. Comparison of AC with state-of-the-art methods on unsupervised segmentation.

Type	Transf.	POS	POS3
None	-	46.4	66.4
Photometric	Col. Jittering	47.9	65.5
Geometric	Flip	46.7	68.0
Geometric	Rot.	**48.5**	68.3
Geo. & Pho.	All	**48.5**	**68.3**

(e) **Transformations:** we apply a given transformation to the inputs of the second forward pass during a single training iteration.

$\lvert\mathcal{O}\rvert$	POS	POS3
2	43.2±2.19	59.5±5.12
4	45.6±3.22	63.55±3.52
8	**46.4**	**66.4**

(b) **Number of orderings:** we compare different sizes of the set $\mathcal{O}$. For $|\mathcal{O}|=2$ and $|\mathcal{O}|=4$, we report the mean and std over 4 runs using different possible pairs and quadruples respectively.

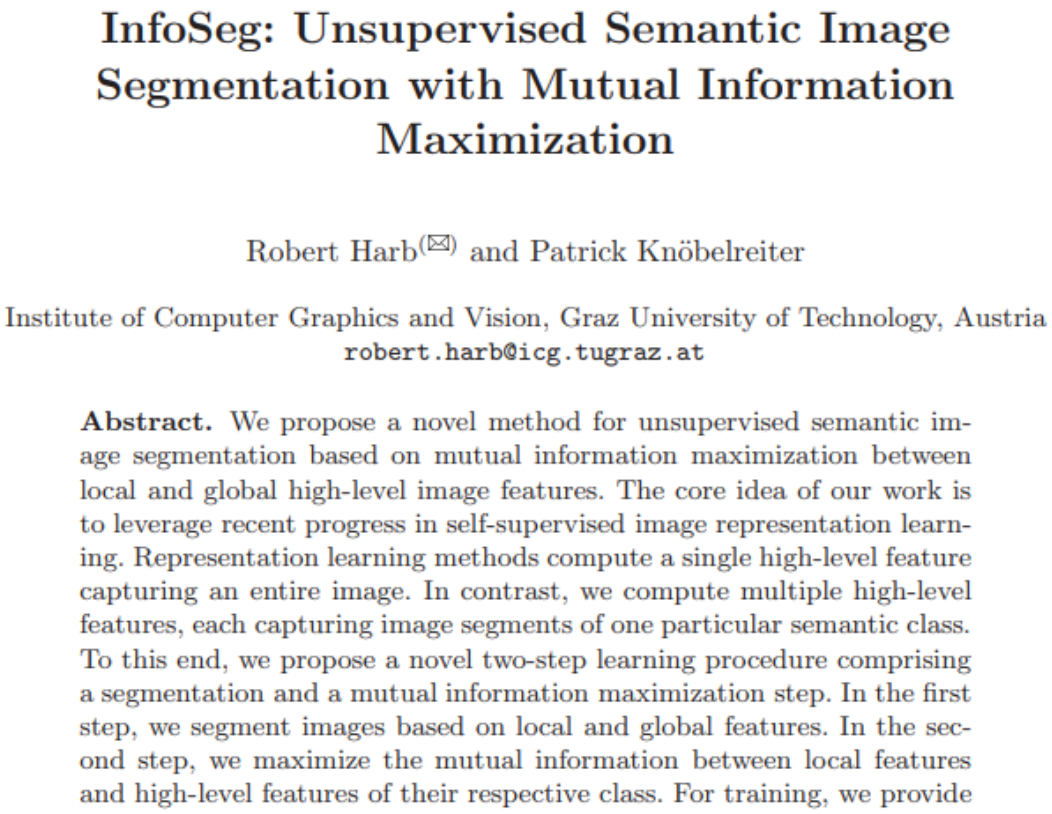


Ouali, Y., Hudelot, C., & Tami, M. (2020, August). Autoregressive unsupervised image segmentation. In European Conference on Computer Vision (pp. 142-158). Springer, Cham.

IV. Methods – 3) InfoSeg

❖ InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization

- Harb *et al.*, German Conference on Pattern Recognition(2021)



InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization

Robert Harb(✉) and Patrick Knöbelreiter

Institute of Computer Graphics and Vision, Graz University of Technology, Austria
robert.harb@icg.tugraz.at

Abstract. We propose a novel method for unsupervised semantic image segmentation based on mutual information maximization between local and global high-level image features. The core idea of our work is to leverage recent progress in self-supervised image representation learning. Representation learning methods compute a single high-level feature capturing an entire image. In contrast, we compute multiple high-level features, each capturing image segments of one particular semantic class. To this end, we propose a novel two-step learning procedure comprising a segmentation and a mutual information maximization step. In the first step, we segment images based on local and global features. In the second step, we maximize the mutual information between local features and high-level features of their respective class. For training, we provide

Unsupervised Semantic Segmentation on COCO-Stuff

Rank	Model	Pixel Accuracy	Paper	Code	Result	Year
1	**InfoSeg**	38.8	InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization			2021
2	**PiCIE**	31.48	PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering			2021
3	**InMARS**	31.0	Unsupervised Image Segmentation by Mutual Information Maximization and Adversarial Regularization			2021
4	**AC**	30.8	Autoregressive Unsupervised Image Segmentation			2020
5	**IIC**	27.7	Invariant Information Clustering for Unsupervised Image Classification and Segmentation			2018

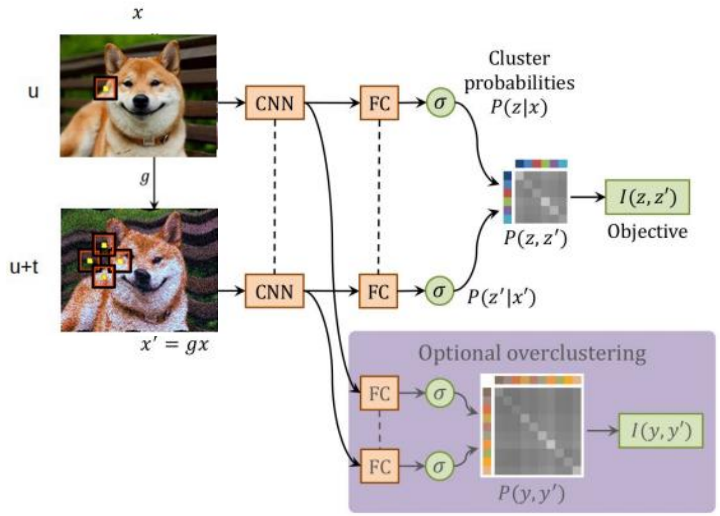
Harb, R., & Knöbelreiter, P. (2021, September). InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization. In DAGM German Conference on Pattern Recognition (pp. 18-32). Springer, Cham.

IV. Methods – 3) InfoSeg

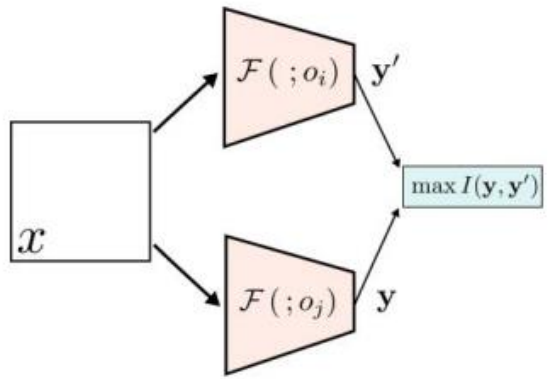
❖ 기존 연구의 한계점 (IIC, AC)

- IIC와 AC 연구에서는 이미지의 **global context**를 모델 학습에 활용하지 않았음을 지적함
- 즉, 픽셀 단위의 클래스 예측은 전체 이미지 정보에 의존적이라고 할 수 있는데, 작은 패치를 통해서만 특징을 추출하는 것은 모델 학습에 충분하지 않음
- **이미지 특징과 segmentation을 동시에 학습**하여 학습 초반에 고차원의 특징 대신 저차원의 특징이 학습됨

IIC (Ji et al., 2019)



AC (Ouali et al., 2020)



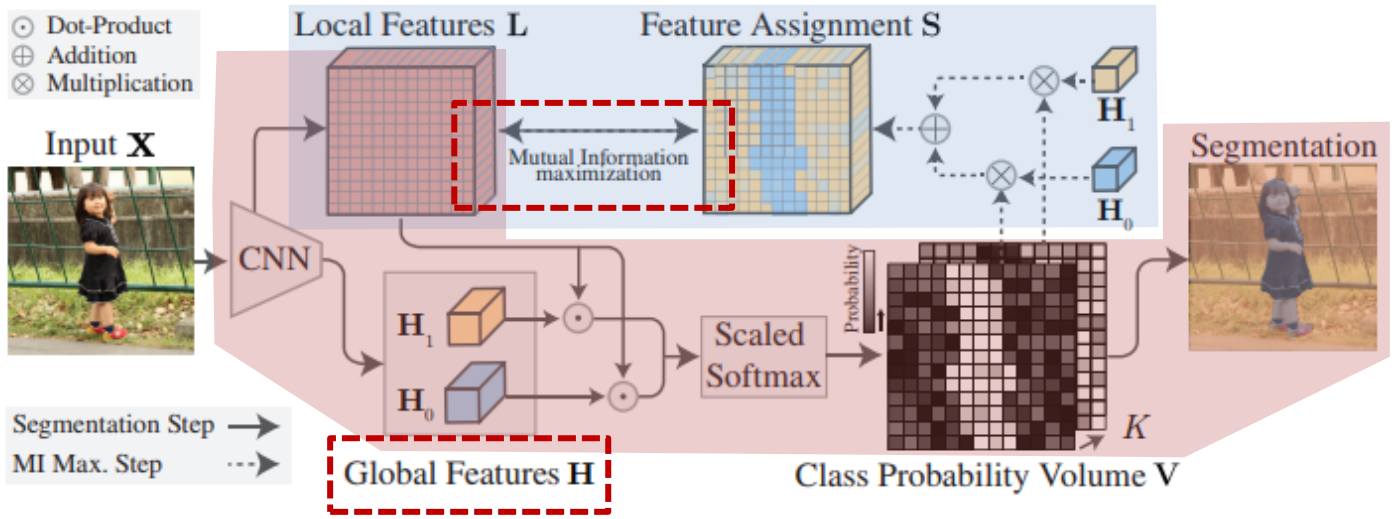
Masked convolution

Harb, R., & Knöbelreiter, P. (2021, September). InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization. In DAGM German Conference on Pattern Recognition (pp. 18-32). Springer, Cham.

IV. Methods – 3) InfoSeg

❖ InfoSeg

- (Segmentation step) 이미지에 대한 local feature와 함께 **global feature**도 학습하며, global feature는 클래스 수만큼 생성
- (Segmentation step) Local feature와 global feature는 다시 결합되고 연산되어 segmentation을 위한 픽셀별 확률을 계산
- (MI Max. step) 픽셀별 확률 분포에 global feature가 다시 적용되어 **local과 global feature의 MI를 최대화**하도록 모델 학습

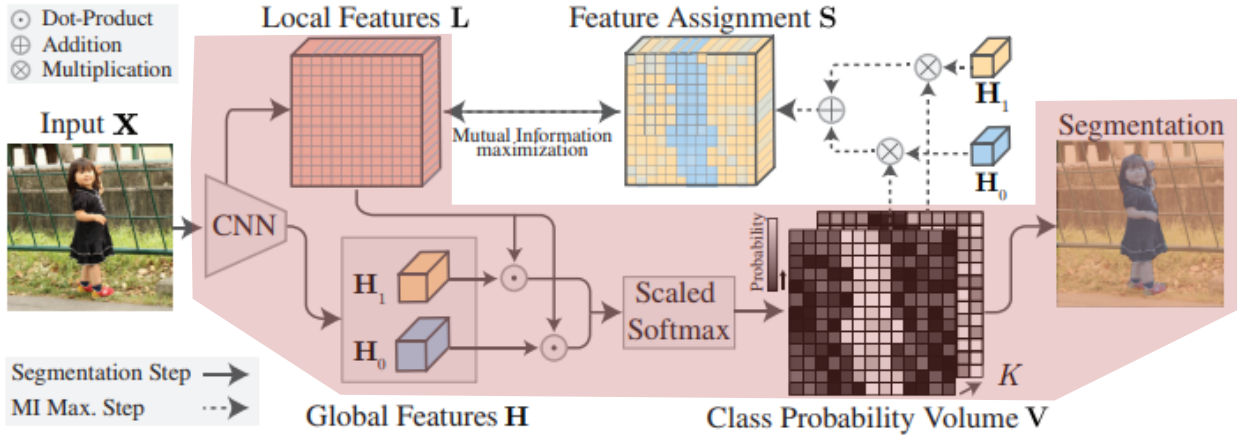


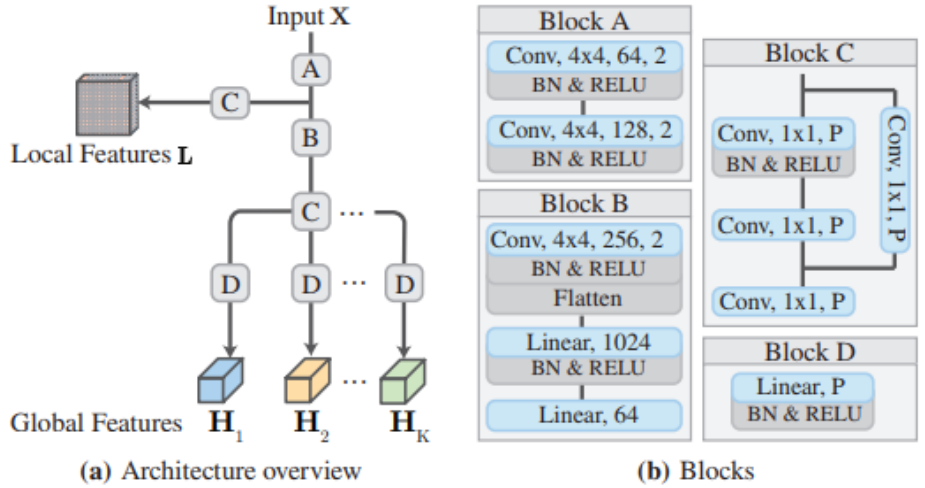
Harb, R., & Knöbelreiter, P. (2021, September). InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization. In DAGM German Conference on Pattern Recognition (pp. 18-32). Springer, Cham.

IV. Methods – 3) InfoSeg

❖ Segmentation step

- **Global feature**는 이미지 전체 영역에 대한 **고차원의 표현을 학습**하며, 목표 클래스 수만큼의 feature를 생성
- **Local feature**는 이미지의 부분 **패치 영역의 특징을 학습**하며 픽셀 위치별로 계산됨
- Local feature와 global feature의 dot product 연산을 통해 픽셀 단위의 클래스 확률을 예측



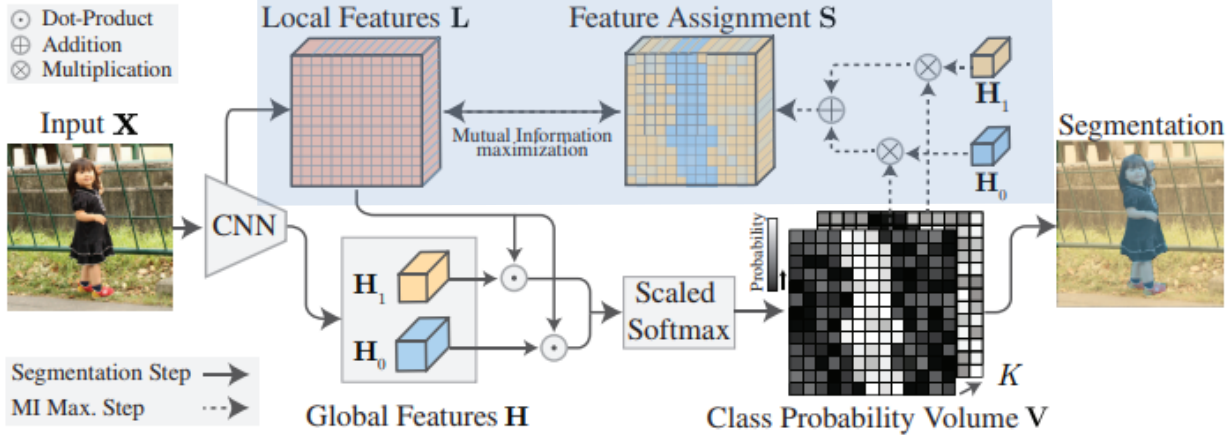


Harb, R., & Knöbelreiter, P. (2021, September). InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization. In DAGM German Conference on Pattern Recognition (pp. 18-32). Springer, Cham.

IV. Methods – 3) InfoSeg

❖ MI maximization step

- **Local feature와 global feature간의 MI를 최대화**하도록 모델을 학습하기 위해 픽셀 단위의 확률 분포(***V***)와 global feature(***H***)를 연산하고 이 결과(***S***)와 local feature(***L***)와의 MI를 최대화



Feature assignment ↓ Class probability ↓ Global feature ↓

$$S_{\theta}^{(i,j)}(\mathbf{X}) = \sum_{k} V_{i,j,k} \cdot H_{\theta}^{(k)}(\mathbf{X}),$$

$$\max_{\theta} \mathbb{E}_{\mathbf{X}} \left[\frac{1}{UV} \sum_{i,j} I\left(L_{\theta}^{(i,j)}(\mathbf{X}); S_{\theta}^{(i,j)}(\mathbf{X}) \right) \right],$$

Mutual information

Harb, R., & Knöbelreiter, P. (2021, September). InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization. In DAGM German Conference on Pattern Recognition (pp. 18-32). Springer, Cham.

IV. Methods – 3) InfoSeg

❖ Experiments

- 이미지의 **local feature와 함께 global feature를 학습**하여 활용함으로써 기존 연구대비 큰 폭의 성능 향상

Method	COCO-Persons	COCO-Stuff	COCO-Stuff-3	Potsdam	Potsdam-3
Random CNN	52.3	19.4	37.3	28.3	38.2
K-Means	54.3	14.1	52.2	35.3	45.7
Doersch* [9]	55.6	23.1	47.5	37.2	49.6
Isola* [16]	57.5	24.3	54.0	44.9	63.9
IIC [17]	57.1	27.7	72.3	45.4	65.1
AC [27]	-	30.8	72.9	49.3	66.5
InMARS [23]	-	31.0	73.1	47.3	70.1
InfoSeg (ours)	**69.6**	**38.8**	**73.8**	**57.3**	**71.6**

Table 1. Pixel-Accuracy of InfoSeg and compared methods. *Clustering of features from methods that are not specifically designed for image segmentation.

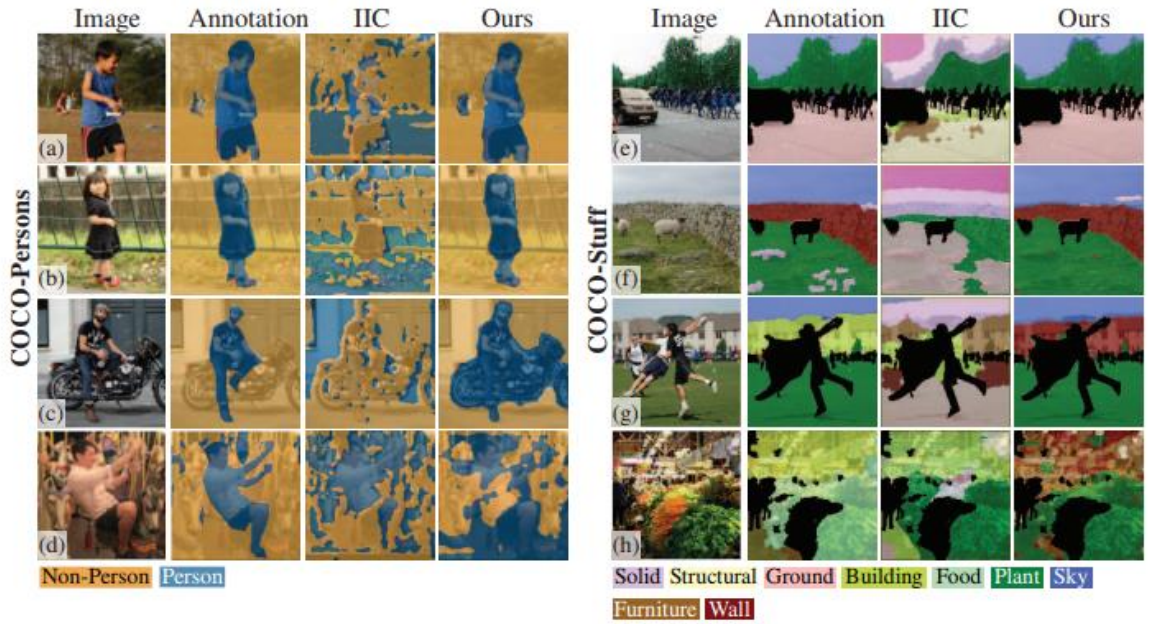


Fig. 4. Qualitative comparison. Non-stuff areas in COCO-Stuff are masked in black.

Harb, R., & Knöbelreiter, P. (2021, September). InfoSeg: Unsupervised Semantic Image Segmentation with Mutual Information Maximization. In DAGM German Conference on Pattern Recognition (pp. 18-32). Springer, Cham.

IV. Methods – 4) PiCIE (Pixel-level feature Clustering using Invariance and Equivariance)

❖ PiCIE: Unsupervised Semantic Segmentation Using Invariance and Equivariance in Clustering

- Cho *et al.* (University of Texas at Austin), 2021 Computer Vision and Pattern Recognition (CVPR)
- 7회 인용 ('22.3.18 기준)

PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering

Jang Hyun Cho[1] Utkarsh Mall[2] Kavita Bala[2] Bharath Hariharan[2]
[1]University of Texas at Austin [2]Cornell University

Figure 1: From these unannotated images, we would like a recognition system to discover the concepts of *house*, *grass*, *trees* and *sky*, and segment each image accordingly without any supervision.

Cho, J. H., Mall, U., Bala, K., & Hariharan, B. (2021). PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16794-16804).

IV. Methods – 4) PiCIE (Pixel-level feature Clustering using Invariance and Equivariance)

❖ PiCIE

- 이미지의 **pixel-level feature representation**과 **k-means clustering**을 동시에 학습하고 이를 통해 픽셀 단위 레이블링
- 입력 이미지에 대한 augmentation을 통해 두 개의 view를 생성하고 각 view에서 얻은 pixel-level feature의 **clustering 결과가 같아지도록** 모델을 학습

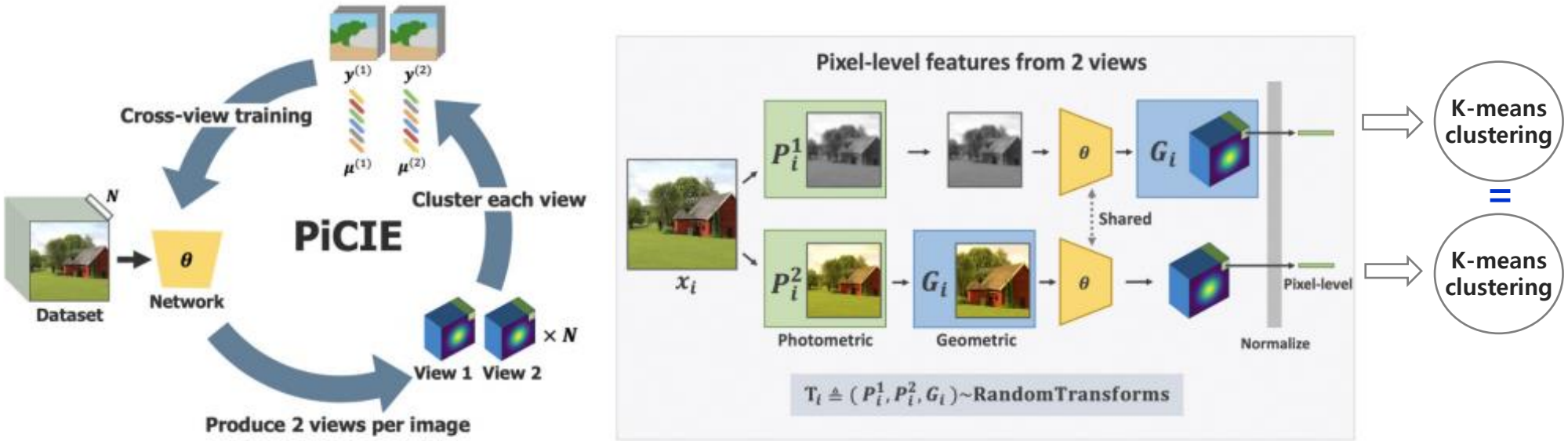


Figure 2: PiCIE overview (**left**) and illustration of multi-view feature computation (**right**). More details in Sec. 3.3.

Cho, J. H., Mall, U., Bala, K., & Hariharan, B. (2021). PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16794-16804).

IV. Methods – 4) PiCIE (Pixel-level feature Clustering using Invariance and Equivariance)

❖ PiCIE - Algorithm

Algorithm 1 PiCIE pseudocode

for $x_i \sim \mathcal{D}$ **do**

$P_i^{(1)}, P_i^{(2)} \sim$ RandomPhotometricTransforms

$G_i \sim$ RandomGeometricTransforms

$z_{i,:}^{(1)} \leftarrow G_i(f_\theta(P_i^{(1)}(x_i)))[:]$

$z_{i,:}^{(2)} \leftarrow f_\theta(G_i(P_i^{(2)}(x_i)))[:]$

end for

$\mu^{(1)}, \mathbf{y}^{(1)} \leftarrow \text{KMeans}(\{z_{ip}^{(1)} : i \in [N], p \in [HW]\})$

$\mu^{(2)}, \mathbf{y}^{(2)} \leftarrow \text{KMeans}(\{z_{ip}^{(1)} : i \in [N], p \in [HW]\})$

for $x_i \sim \mathcal{D}$ **do**

$z_{i,:}^{(1)} \leftarrow G_i(f_\theta(P_i^{(1)}(x_i)))[:]$

$z_{i,:}^{(2)} \leftarrow f_\theta(G_i(P_i^{(2)}(x_i)))[:]$

$\mathcal{L}_{\text{within}} \leftarrow \sum_p \mathcal{L}_{\text{clust}}(z_{ip}^{(1)}, y_{ip}^{(1)}, \mu^{(1)}) + \mathcal{L}_{\text{clust}}(z_{ip}^{(2)}, y_{ip}^{(2)}, \mu^{(2)})$

$\mathcal{L}_{\text{cross}} \leftarrow \sum_p \mathcal{L}_{\text{clust}}(z_i^{(1)}, y_{ip}^{(2)}, \mu^{(2)}) + \mathcal{L}_{\text{clust}}(z_{ip}^{(2)}, y_{ip}^{(1)}, \mu^{(1)})$

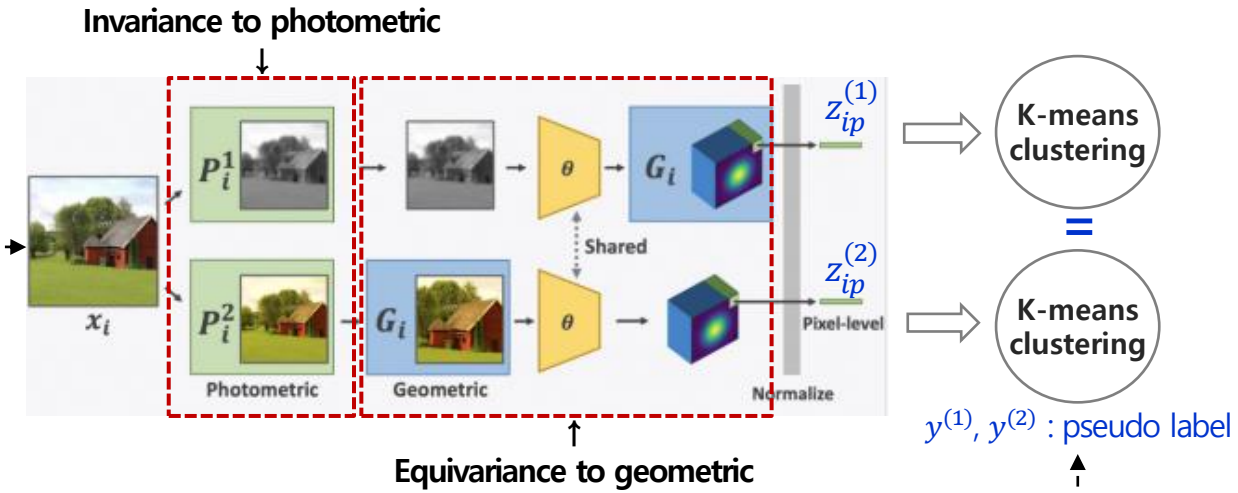
$\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{within}} + \mathcal{L}_{\text{cross}}$

$f_\theta \leftarrow \text{backward}(\mathcal{L}_{\text{total}})$

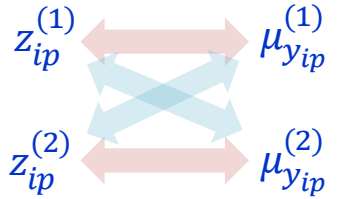
end for

Above: PiCIE pseudo-code. Notations consistent with Sec. 3.3.

Photometric transformations : color jitter, gray scale, gaussian blur
Geometric transformations : random crop, horizontal flip



$$\mathcal{L}_{clust}(f_\theta(x_i)[p], y_{ip}, \boldsymbol{\mu}) = -\log\left(\frac{e^{-d(f_\theta(x_i)[p], \mu_{y_{ip}})}}{\sum_l e^{-d(f_\theta(x_i)[p], \mu_l)}}\right)$$



Cho, J. H., Mall, U., Bala, K., & Hariharan, B. (2021). PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16794-16804).

IV. Methods – 4) PiCIE (Pixel-level feature Clustering using Invariance and Equivariance)

❖ Experiments

COCO-*All* : 27 classes

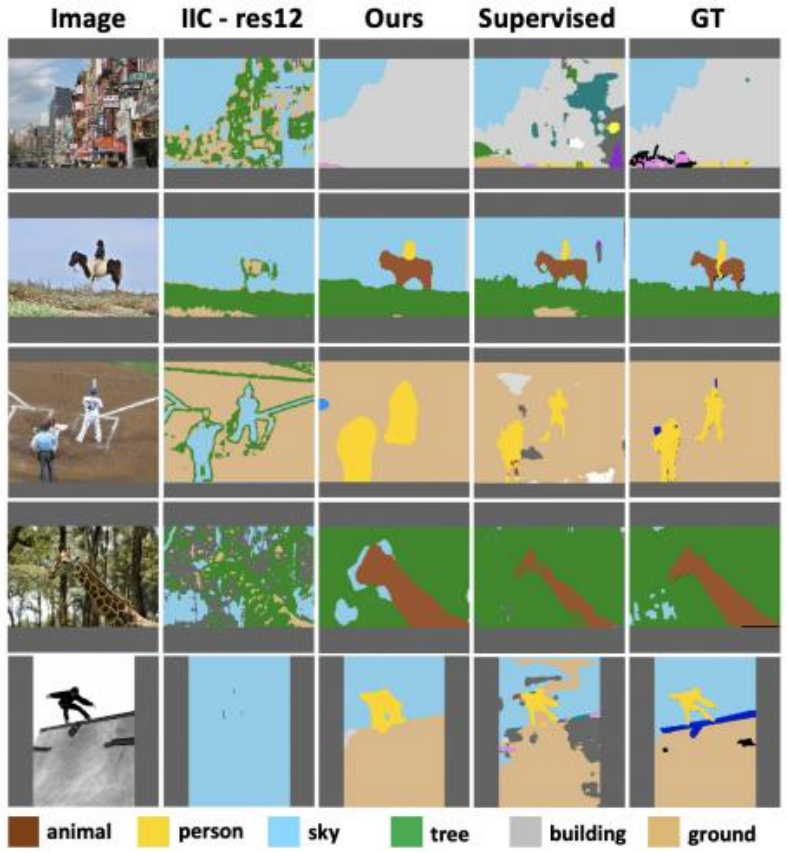
Method	Classifier	Acc.	mIoU
No Train	Linear	17.45	3.70
No Train	Prototype	26.26	8.41
Modified DC	Linear	32.21	9.79
IIC - res12 [23]	Linear	22.45	4.11
IIC [23]	Linear	21.79	6.71
PiCIE	Prototype	48.09	13.84
PiCIE + H.	Prototype	**49.99**	**14.36**

Table 1: COCO-*All* [23] results. Our method is compared to clustering methods adapted to semantic segmentation. "+H." denotes PiCIE trained with auxiliary clustering.

COCO-*Stuff* : 15 classes

Method	COCO-*Stuff*
Random CNN	19.4
K-means [38]	14.1
SIFT [33]	20.2
Doersch 2015 [10]	23.1
Isola 2016 [21]	24.3
DeepCluster [4]	19.9
IIC [23]	27.7
AC [37]	30.8
Modified DC	25.26
IIC	27.97
IIC – res12	27.92
PiCIE	**31.48**

Table 4: COCO-*Stuff* results without ImageNet pretrained weight following [23, 37]. First section is from prior works [23, 37] and the last two sections are from our implementation.



Cho, J. H., Mall, U., Bala, K., & Hariharan, B. (2021). PiCIE: Unsupervised Semantic Segmentation using Invariance and Equivariance in Clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16794-16804).

I. Introduction

II. Unsupervised semantic segmentation

III. Mutual information maximization

IV. Methods

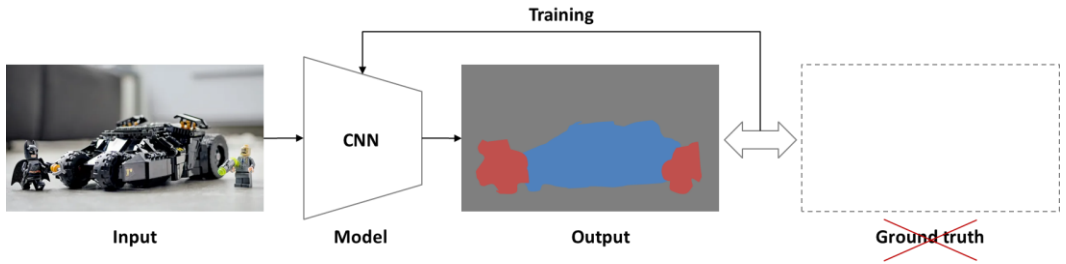
V. Conclusion

V. Conclusion

❖ 결론

- Semantic segmentation 문제는 픽셀 단위의 레이블을 얻기 어렵기 때문에 **입력 이미지만으로 모델을 학습**할 수 있는 Unsupervised learning 방법론 연구가 필요
- Unsupervised semantic segmentation 연구는 입력 이미지의 Augmentation을 통해 **Positive pair**를 생성하고 이를 모델에 통과시킨 후, **Mutual information maximization**을 통해 모델을 학습시키는 방법으로 최근 연구되고 있음
- Mutual information은 상호의존성을 측정하여 **두 변수간의 독립성을 판단**할 수 있고, 이를 Unsupervised semantic segmentation에서는 Positive pair간의 **공통 특징 추출**과 클러스터링에서의 **Degenerate solution을 해결**하기 위해 사용됨
- IIC, AC, InfoSeg, PiCIE 등의 최근 연구들은 Positive pair와 Mutual information을 활용하여 성능을 향상시키고 있으나, 아직 **개선 여지가 많고 다양한 방향의 연구가 필요**

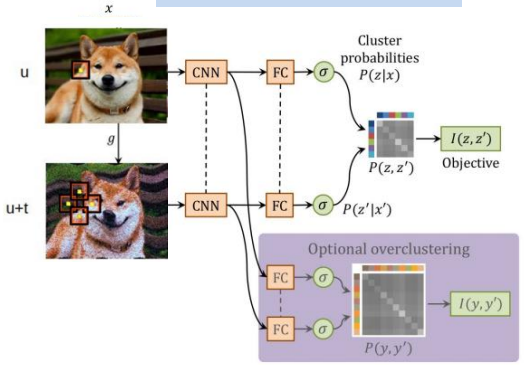
Unsupervised semantic segmentation



Mutual information

$$I(X;Y) \triangleq D_{KL}\big(P(x,y) \parallel P(x)P(y)\big)$$
$$= \sum_{y\in Y}\sum_{x\in X} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}$$
$$= H(X) + H(Y) - H(X,Y)$$
$$= H(X) - H(X|Y)$$
$$= H(Y) - H(Y|X)$$

IIC(Ji *et al.*, 2019)



감사합니다.