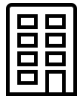
『2025 겨울 세미나』

OoD Segmentation



Pusan National University
Dept. of Electronics Engineering
Computer Vision & Signal Processing Lab.



Presented By
우민수

Outline

- Background
- RbA: Segmenting Unknown Regions Rejected by All [ICCV 23]
- Diffusion for Out-of-Distribution Detection on Road Scenes and Beyond[ECCV 24]

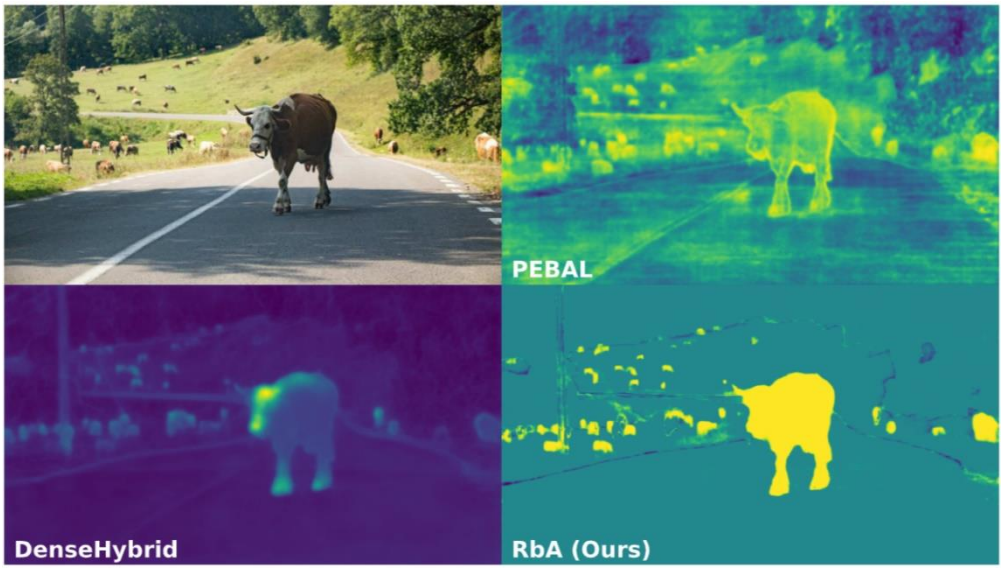
Background

- OoD Segmentation
  - Open-Set Semantic Segmentation
    - 일반적인 Segmentation이 closed-set class를 가정으로 하는 것과는 달리 새로운 클래스도 처리 가능
      - 예를 들어 학습 클래스에 없는 사슴이 도로에 나타났을 때 이를 탐지
  - Anomaly Segmentation
    - 이미지 내에서 정상(inlier)와 비정상(outlier)를 구별하는 것

RbA: Segmenting Unknown Regions Rejected by All [ICCV 23]

Introduction

- Task
  - Detecting OoD(Out of Distribution)
    - 머신러닝 모델이 학습한 분포(In-Distribution)에서 벗어난 새로운 데이터(Out of Distribution)를 효과적으로 탐지하는 것



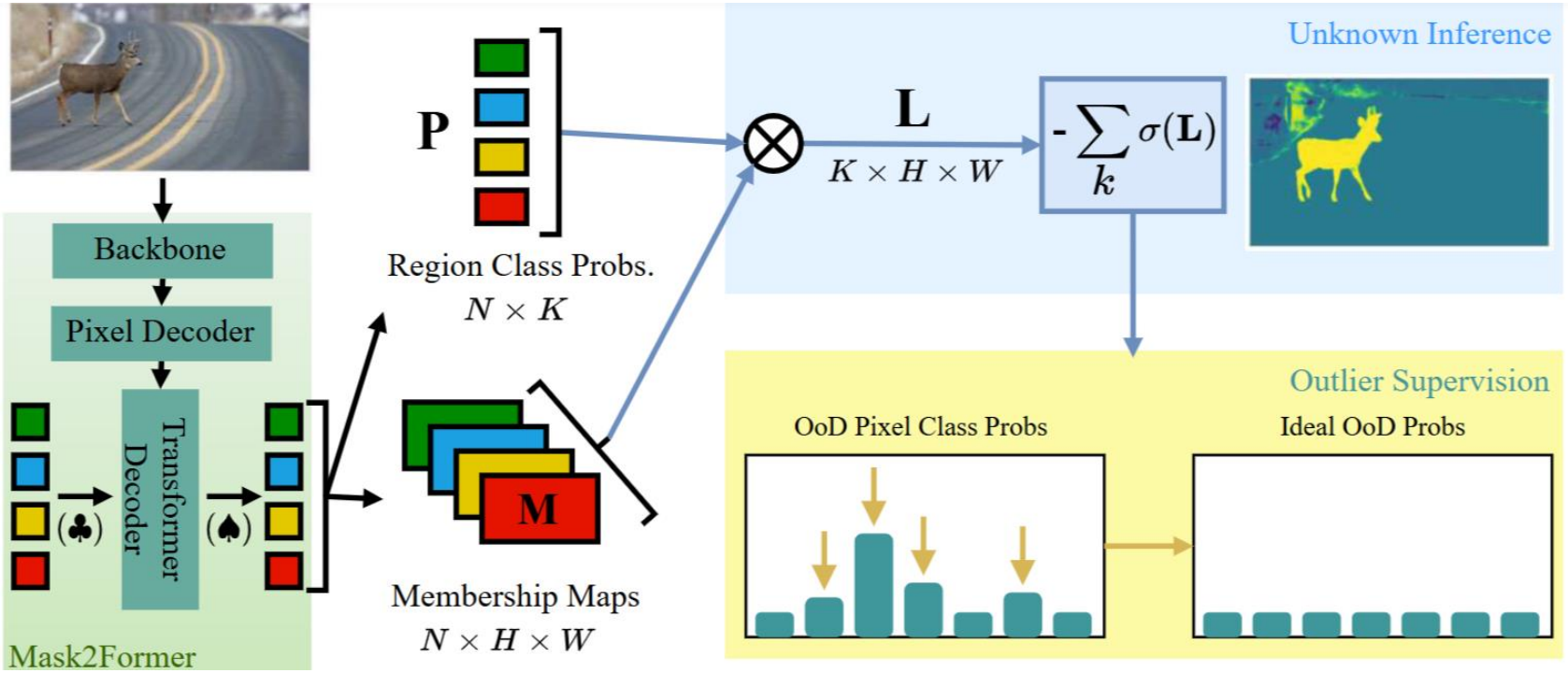
< Result of RbA >

Introduction

- Problem & Solution
  - Existing Image Segmentation models
    - 픽셀별로 일일이 클래스를 예측해 두 물체사이 경계의 경우 픽셀들이 두 물체의 클래스 중 어디에 속하는지 헷갈려 하고 예측 확률이 낮게 나와서 outlier로 판단하는 경우가 발생
  - Proposed Method
    - Image Segmentation를 위한 새로운 방식을 제안

Method

- RbA
  - Total

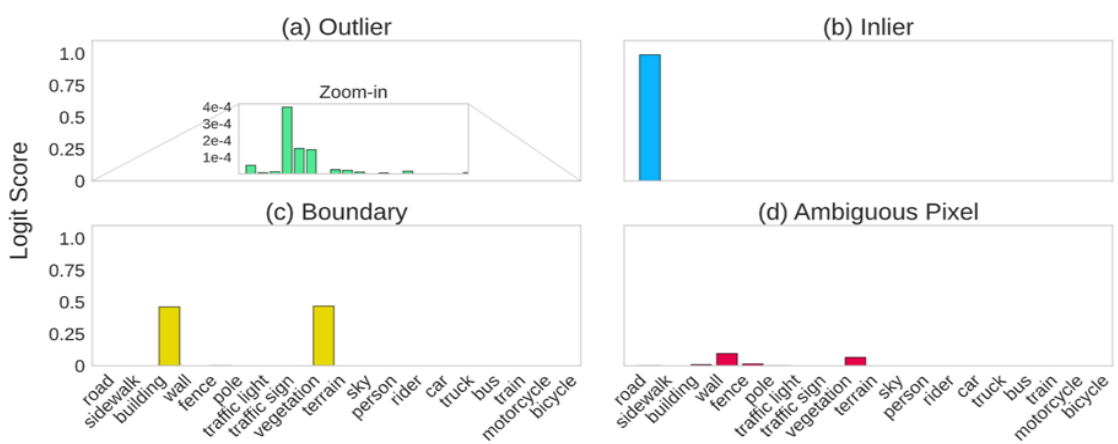


< Overview of RbA>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- Overview
  - Mask2Former
    - 픽셀별로 일일이 어떤 클래스에 속하는지 예상하는 것에서 벗어나 먼저 특정 영역이 어디에 위치하는지를 예측한 후 해당영역이 어떤 클래스에 속하는지 구별
  - RbA Score
    - 기존의 Softmax 기반의 클래스 확률 계산 및 지정 대신 outlier를 걸러내기 위해 One vs All Classifier 방식으로 구한 모든 클래스에 대한 확률을 더하여 –를 붙여줌

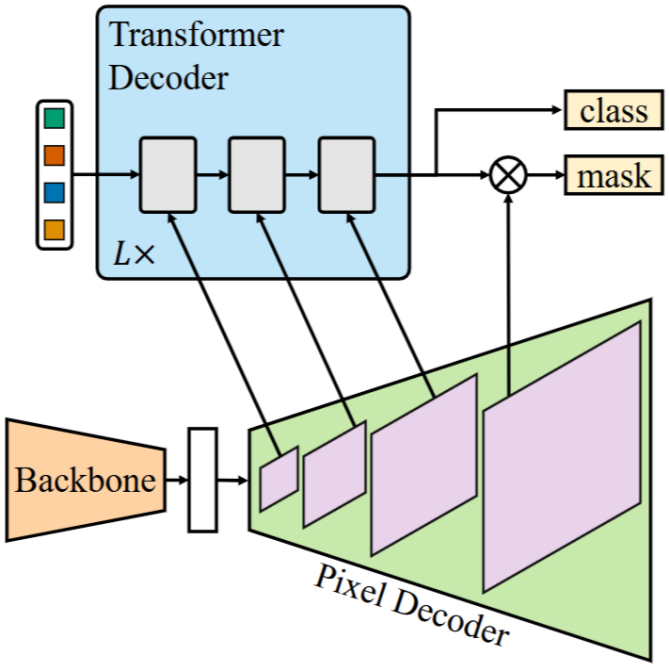


< Overview of RbA Score>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- RbA
  - Mask2Former
    - Backbone, Pixel Decoder, Transformer Decoder
      - Backbone:이미지에서 feature를 추출
      - Pixel Decoder:Backbone에서 추출된 feature를 고해상도 feature로 처리
      - Transformer Decoder:Pixel Decoder에서 추출된 고해상도 feature와 object query를 이용해 클래스와 영역을 예측

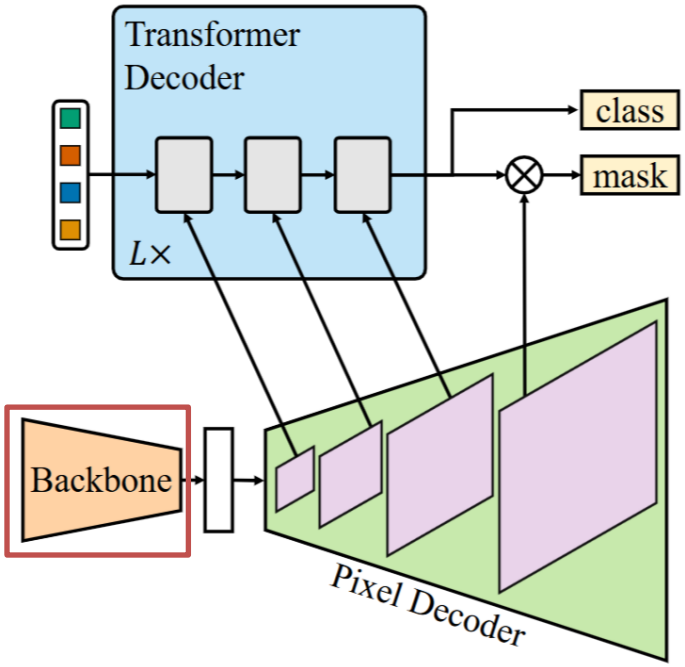


< Overview of Mask2Former>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- RbA
  - Mask2Former
    - Backbone, Pixel Decoder, Transformer Decoder
      - Backbone:이미지에서 feature를 추출
      - Pixel Decoder:Backbone에서 추출된 feature를 고해상도 feature로 처리
      - Transformer Decoder:Pixel Decoder에서 추출된 고해상도 feature와 object query를 이용해 클래스와 영역을 예측

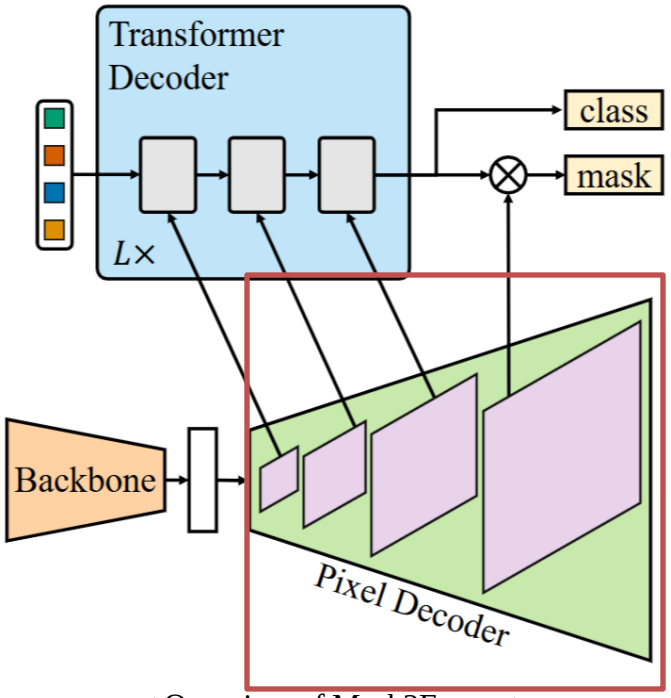


< Overview of Mask2Former>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

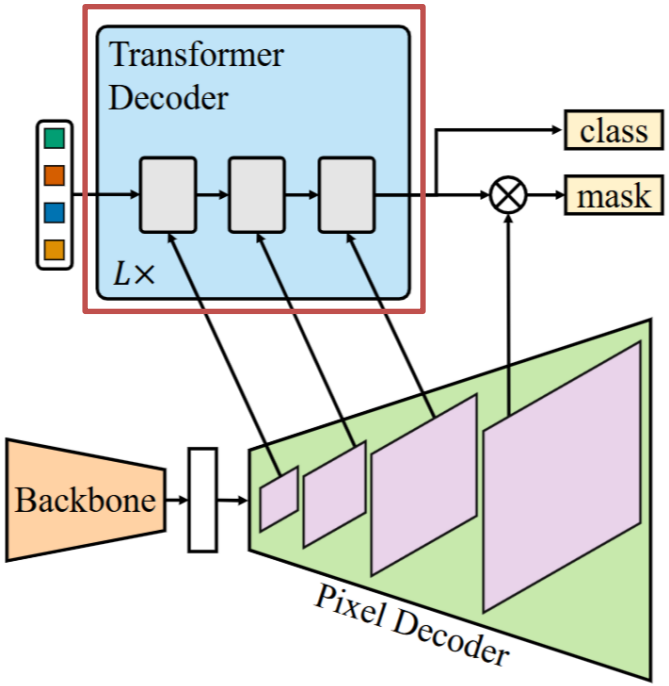
- RbA
  - Mask2Former
    - Backbone, Pixel Decoder, Transformer Decoder
      - Backbone:이미지에서 feature를 추출
      - Pixel Decoder:Backbone에서 추출된 feature를 고해상도 feature로 처리
      - Transformer Decoder:Pixel Decoder에서 추출된 고해상도 feature와 object query를 이용해 클래스와 영역을 예측



< Overview of Mask2Former>

Method

- RbA
  - Mask2Former
    - Backbone, Pixel Decoder, Transformer Decoder
      - Backbone:이미지에서 feature를 추출
      - Pixel Decoder:Backbone에서 추출된 feature를 고해상도 feature로 처리
      - Transformer Decoder:Pixel Decoder에서 추출된 고해상도 feature와 object query를 이용해 클래스와 영역을 예측



< Overview of Mask2Former>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- RbA
  - Mask Classification
    - Object query를 이용해 특정 영역의 범위를 정하고 동시에 해당 영역의 클래스를 결정

$$M(x) = \sigma(Q_p \cdot F(x))$$

Membership Map: 각 픽셀에 대해 객체가 속할 가능성이 있는 마스크 확률 맵으로서 object query에서 뽑아낸 특징 벡터 Q_p와 pixel decoder에서 뽑아낸 고해상도 feature map을 내적하고 시그모이드 함수를 통과시켜 완성 object query의 영역을 결정하는데 사용됨

$$P(x) \in [0,1]^{N \times K}$$

각 object query(N개)가 예측된 영역에서 어떤 클래스(K개)에 속할지에 대한 확률을 구하는 식으로서 transformer decoder에서 에측됨

$$L(x) = \sum_{n=1}^{N} P_n(x) M_n(x)$$

Class logit: 각 객체 쿼리의 클래스 확률 $Pn(x)$과 마스크 $Mn(x)$을 가중 평균하여 최종 클래스 로짓을 얻음 즉, 각 픽셀에 대해 "모든 객체 쿼리의 예측을 반영"하여 최종적으로 어느 클래스인지 결정

Method

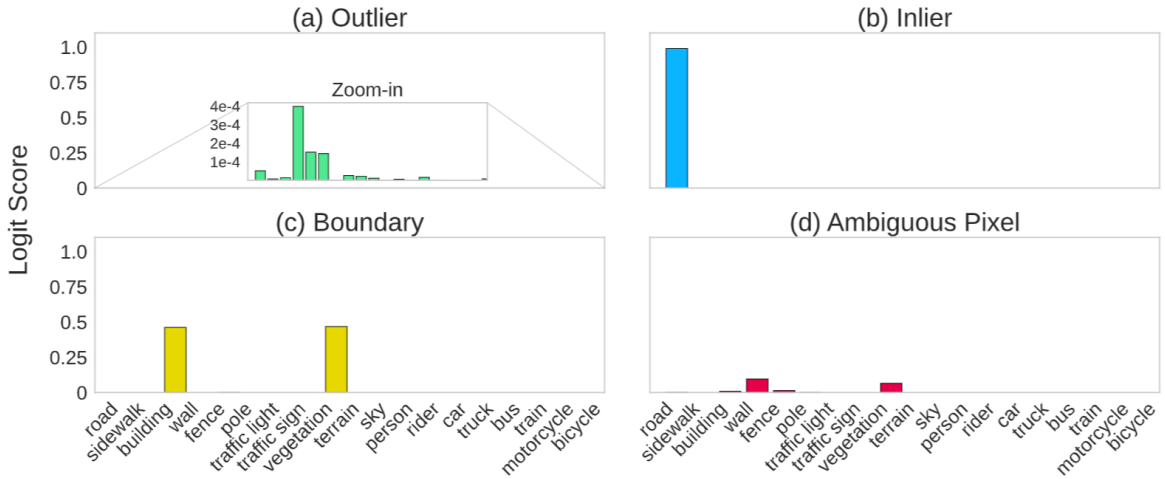
- RbA
  - RbA Score

– 각 클래스 별 확률을 모두 더하고 –를 붙인 값, 커질수록 outlier일 확률이 높아짐

$$p(\mathbf{y} = K+1|\mathbf{x}) = 1 - \sum_{k=1}^{K} p(\mathbf{y} = k|\mathbf{x})$$

$$= 1 - \sum_{k=1}^{K} \sigma(\mathbf{L}_k(\mathbf{x}))$$

$$\mathrm{RbA}(\mathbf{x}) = -\sum_{k=1}^{K} \sigma(\mathbf{L}_k(\mathbf{x}))$$



< Overview of RbA score>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- RbA
  - Fine tuning
    - closed-form에서의 성능 저하를 막으면서 outlier 검출성능도 향상
      - 전체 모델중 일부(0.21%)만 미세 조정하여 위 목적을 달성
      - 합성 이상 데이터셋으로 위 모델 중 일부를 Fine Tuning
      - RbA값이 충분히 커지도록 학습을 유도
      - $$L_{RbA} = \sum_{x \in \Omega_{\text{out}}} \left(\max(0, \alpha - RbA(x))\right)^2$$

Experiments

- Quantiative Results
  - SMIYC Dataset
    - Metric
      - AP, FPR, sIoU gt, PPV, mean F1

Method	OoD Data	Extra Net.	Anomaly Track					Obstacle Track				
			AP ↑	FPR ↓	sIoU gt ↑	PPV ↑	mean F1 ↑	AP ↑	FPR ↓	sIoU gt ↑	PPV ↑	mean F1 ↑
Emb. Density[5]	✗	✗	37.5	70.8	33.9	20.5	7.9	0.8	46.4	35.6	2.9	2.3
JSRNet[66]	✗	✓	33.6	43.9	20.2	29.3	13.7	28.1	28.9	18.6	24.5	11.0
Road Inpain.[47]	✗	✓	-	-	-	-	-	54.1	47.1	**57.6**	39.5	36.0
Image Resyn.[48]	✗	✓	52.3	25.9	39.7	11.0	12.5	37.7	4.7	16.6	20.5	8.4
ObsNet[1]	✗	✓	75.4	26.7	44.2	**52.6**	**45.1**	-	-	-	-	-
NFlowJS[24]	✗	✓	56.9	34.7	36.9	18.0	14.9	85.6	**0.4**	45.5	49.5	**50.4**
RbA (Ours)	✗	✗	**86.1**	**15.9**	**56.3**	41.4	42.0	**87.8**	3.3	47.4	**56.2**	**50.4**
Max. Entropy[43]	✓	✗	85.5	15.0	49.2	39.5	28.7	85.1	0.8	47.9	**62.6**	48.5
DenseHybrid[25]	✓	✗	78.0	**9.8**	54.2	24.1	31.1	87.1	**0.2**	45.7	50.1	50.7
PEBAL[65]	✓	✗	49.1	40.8	38.9	27.2	14.5	5.0	12.7	29.9	7.6	5.5
SynBoost[4]	✓	✓	56.4	61.9	34.7	17.8	10.0	71.3	3.2	44.3	41.8	37.6
RbA (Ours)	✓	✗	**90.9**	11.6	**55.7**	**52.1**	**46.8**	**91.8**	0.5	**58.4**	58.8	**60.9**

< Quantitative comparisons on the SMIYC Dataset>

Experiments

- Quantiative Results
  - SMIYC Dataset
    - Metric
      - AP

        Precision-Recall 곡선의 **면적으로서, 탐지의 정확성과 재현율의 균형을 평가하는 지표**

        AP 값이 높을수록, **모델이 올바르게 이상치를 감지할 확률이 높음**
      - FPR

        실제 정상(Inlier)인 샘플을 이상치(Outlier)로 잘못 분류한 비율로서, **잘못된 이상치 탐지(False Alarm)가 얼마나 발생했는지** 평가하는 지표, FPR이 **낮을수록**, 정상 데이터를 이상치로 잘못 판단하는 빈도가 줄어듦
      - sIoU gt

        모델이 예측한 이상치 영역과 **실제 이상치 영역(GT) 간의 겹치는 정도**를 평가

        1에 가까울수록, 예측된 이상치 영역이 실제 GT와 잘 맞아떨어짐

Experiments

- Quantiative Results
  - SMIYC Dataset
    - Metric
      - PPV

        모델이 **이상치라고 예측한 픽셀 중, 실제로 이상치인 픽셀 비율**

        PPV가 **높을수록**, 모델이 **이상치 탐지를 정확하게 수행**했다는 의미
      - mean F1

        Precision(정밀도)과 Recall(재현율)의 조화평균(Harmonic Mean)

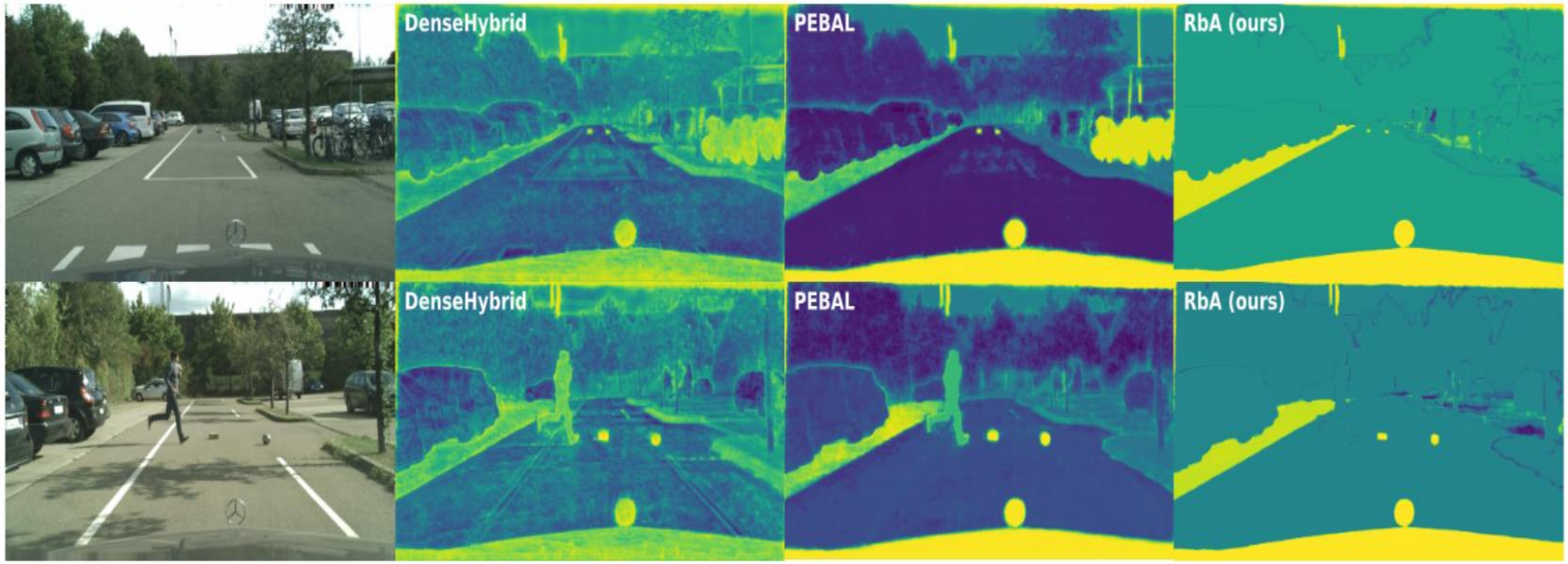
        Precision(오탐 방지)과 Recall(정탐 성능)을 동시에 고려하여 최적의 성능을 평가

        1에 가까울수록 좋은 성능

Experiments

- Qualitative Results

PEBAL과 DenseHybrid 모두 사람을 outlier로 판단하는 것과 대비하여 RbA는 사람을 outlier로 판단하지 않는 것을 알 수 있음



< Qualitative comparisons on the SMIYC Dataset >

- **Diffusion for Out-of-Distribution Detection on Road Scenes and Beyond[ECCV 24]**

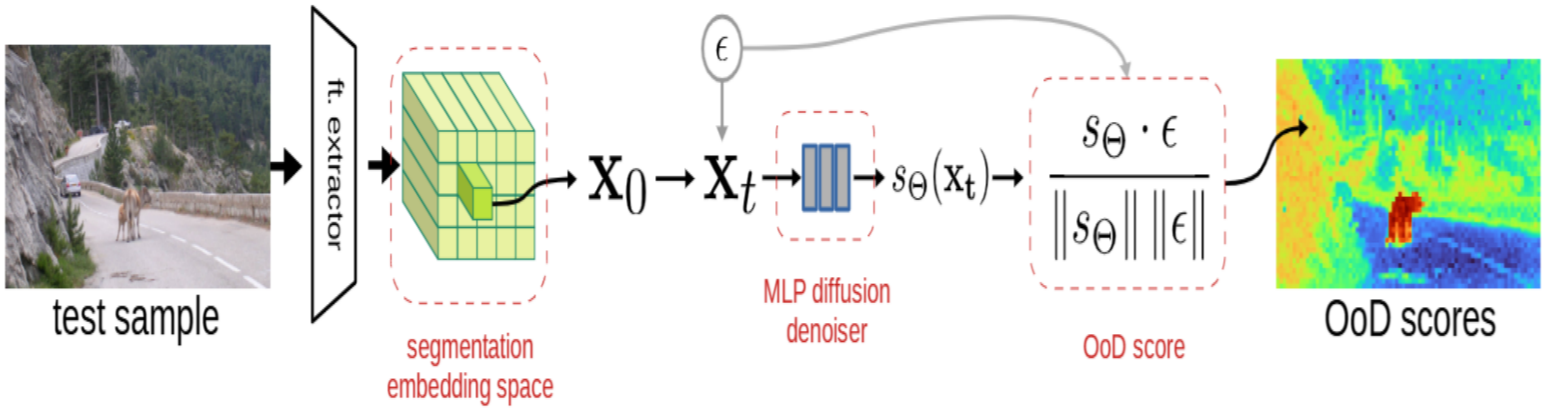
Introduction

- Problem & Solution
  - Existing Image Segmentation models
    - 기존 OoD 탐지 연구는 주로 도로 장면(Road Scenes)을 대상으로 했으며, 특정 도메인에 한정됨, 또한 기존 방법들은 **OoD 탐지를 위해 추가적인 이상 데이터(OoD)를 필요로 하거나, 특정 데이터 도메인에 의존하는 한계**가 존재
  - Proposed Method
    - ADE20K 데이터셋을 기반으로 다양한 자연 이미지에서 OoD 탐지를 수행할 수 있도록 확장, **In-Distribution 데이터만을 사용하여 확률 분포를 학습하고, Score Function을 활용해 OoD를 탐지하는 새로운 접근법(DOoD)을 제안**

Method

- Overview
  - DOoD(Diffusion Score Matching for OoD Detection)
    - DOoD는 Semantic Segmentation 모델에서 추출한 특징을 Diffusion Process에 적용하고, Score Function을 학습하여 OoD를 효과적으로 탐지하는 방법

      Road scene에만 최적화 되어있고 느렸던 기존 방법에 비해 다양한 scene에 적용가능하며 빠름



< Overview of DOoD >

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

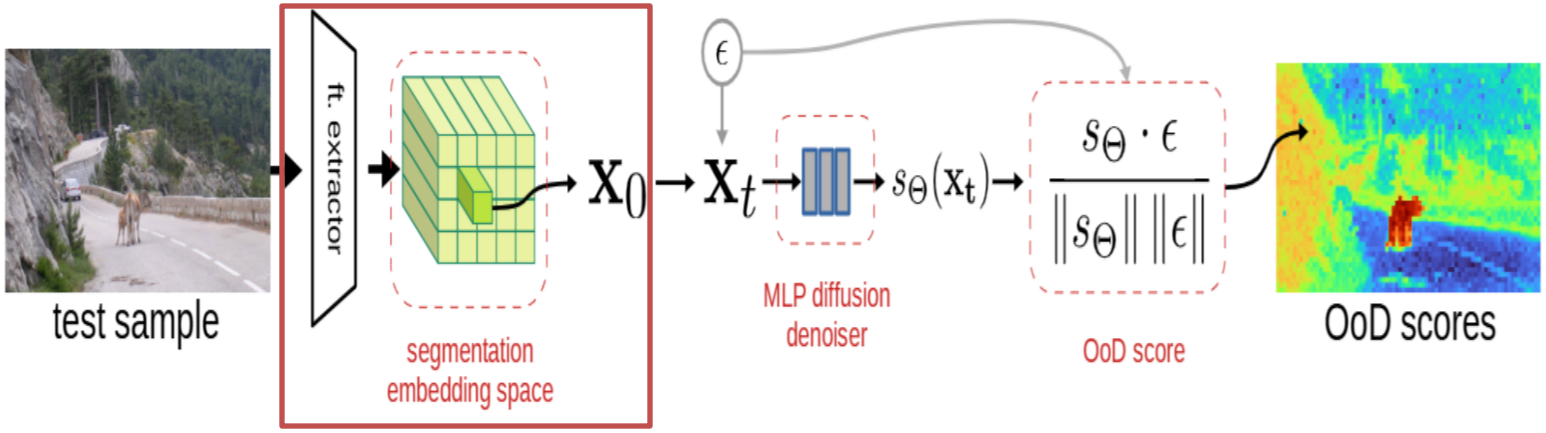
- Overview
  - ADE-OoD
    - 기존의 벤치마크 데이터셋에 비해 훨씬 다양한 클래스 포함
      - 기존의 벤치마크들이 19개 이하의 클래스를 가지는 반면에 ADE-OoD는 150개의 클래스를 가짐

	Road-Anomaly	Fishyscapes L&F	Fishyscapes Static	SMIYC Anomaly	SMIYC Obstacle	ADE-OoD (Ours)
Type	Real	Real	Synthetic (pasted cutouts)	Real	Real (road only)	Real
Domain	Driving	Driving	Driving	Driving	Driving	In & outdoor
Num. Samples	60	100+275	30+1000	10+100	327+30	111
In-dist. classes	19	19	19	19	1	150

< Overview of ADE-OoD >

Method

- DOoD
  - Feature extracting
    - 입력 이미지를 Pre-trained Image Segmentation Model의 Encoder(ViT 또는 MiT)를 통과하여 Feature Map X_o을 생성

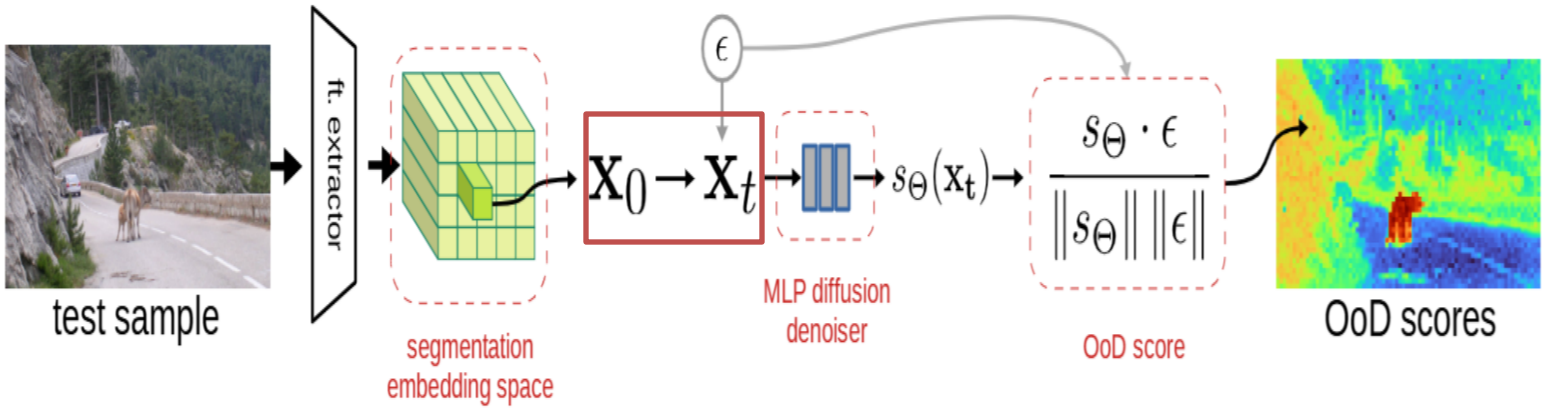


< Overview of DOoD>

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- DOoD
  - Forward diffusion process
    - 추출된 Feature Map X_o에 가우시안 노이즈 ϵ를 추가하여 점진적으로 변형된 X_t를 생성

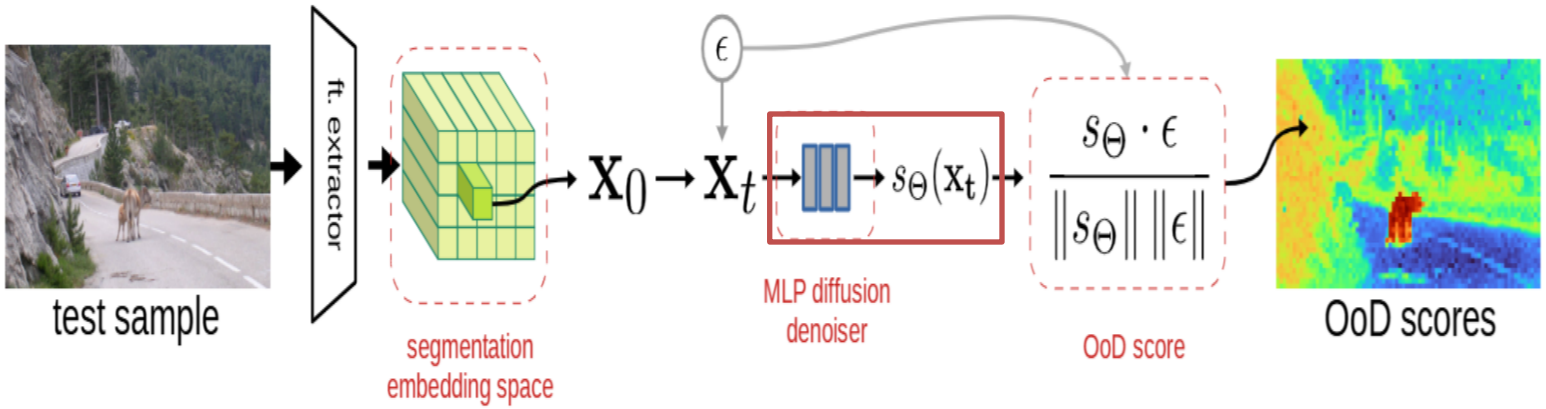


< Overview of DOoD >

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- DOoD
  - Reverse diffusion process
    - 노이즈가 추가된 특징 X_t 를 MLP 기반 denoiser를 통과하여 score function을 예측
      - 기존 U-net기반 denoiser와 달리, MLP 구조를 사용해 공간적 상관관계를 줄임



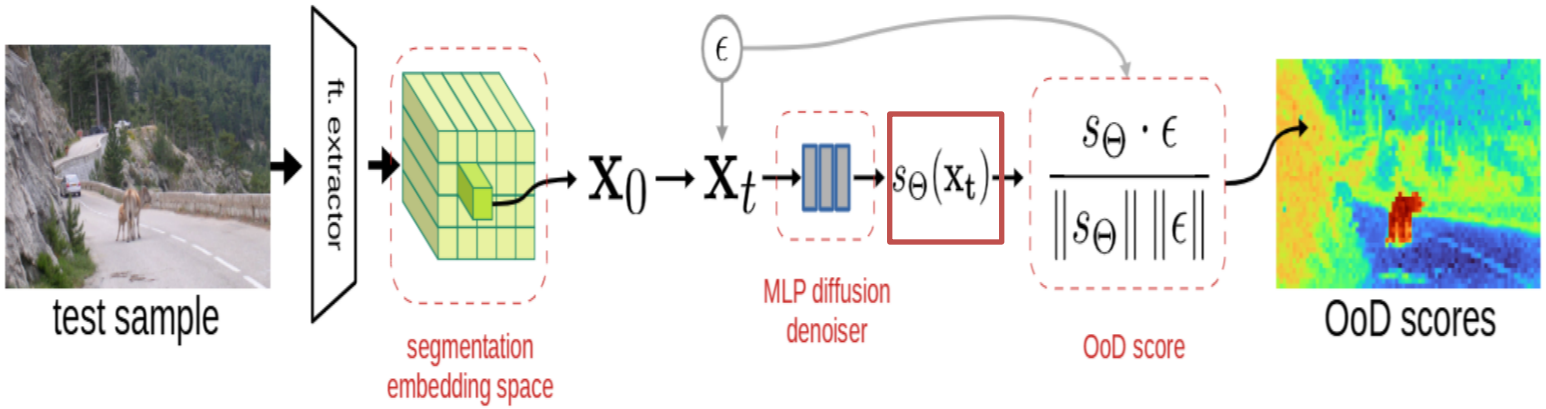
< Overview of DOoD >

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- DOoD
  - Score function
    - 로그 확률 밀도함수의 기울기로써 어떤 방향으로 이동해야 정상 데이터 분포에 가까워지는지를 알려줌, 이를 구하는 것은 실질적으로 노이즈를 제거하는 것과 같음

$$s_\Theta(X_t) = \nabla_{X_t} \log p(X_t)$$

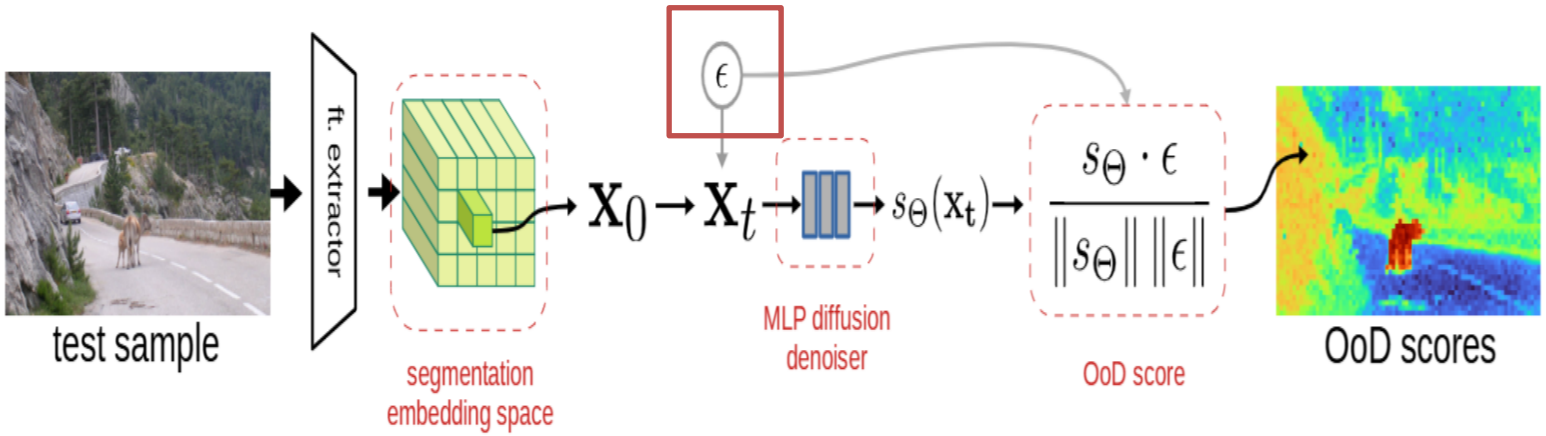


< Overview of DOoD >

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- DOoD
  - Noise ϵ
    - 정상 데이터 분포에서 멀어지거나 가까워지는 방향으로 더해지거나 제거되는 noise로써 ideal한 경우 inlier 이미지가 들어와서 구해지는 score function과 평행한 방향의 벡터



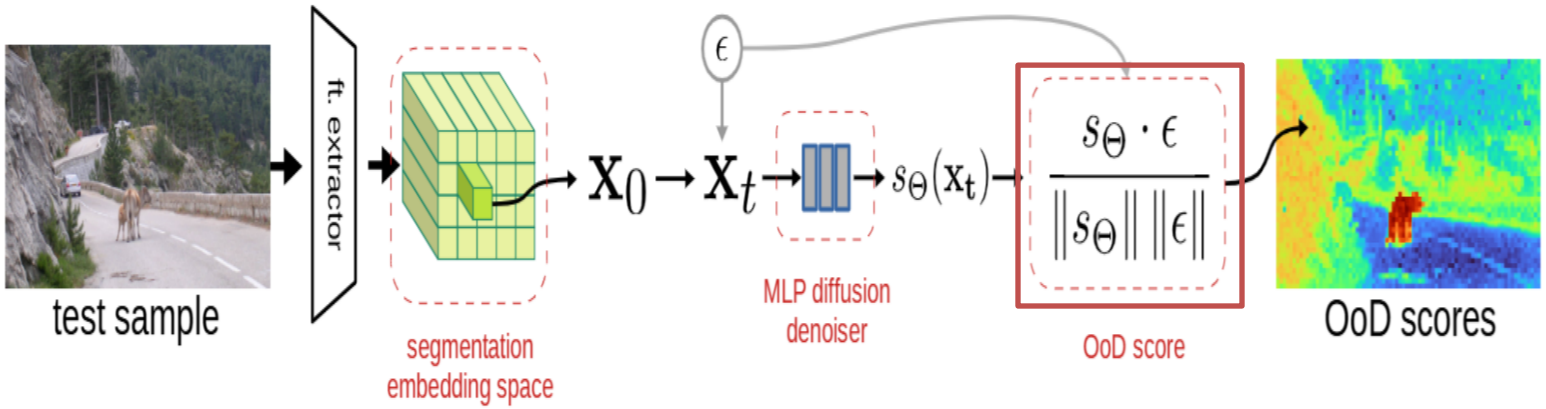
< Overview of DOoD >

부산대학교 PUSAN NATIONAL UNIVERSITY
CVSP Lab Computer Vision & Signal Processing

Method

- DOoD
  - OoD Score
    - 노이즈 ϵ 와 score function의 cosine similarity로써 이 값이 크면 정상 데이터(In-Distribution)에 가깝고 이 값이 작으면 OoD 데이터(이상 데이터)일 가능성이 높음

$$e_{\angle} = \frac{s_\Theta(X_t) \cdot \epsilon}{\|s_\Theta(X_t)\| \|\epsilon\|}$$



< Overview of DOoD >

Experiments

- Quantiative Results
  - ADE-OoD Dataset
    - Metric
      - AP, FPR

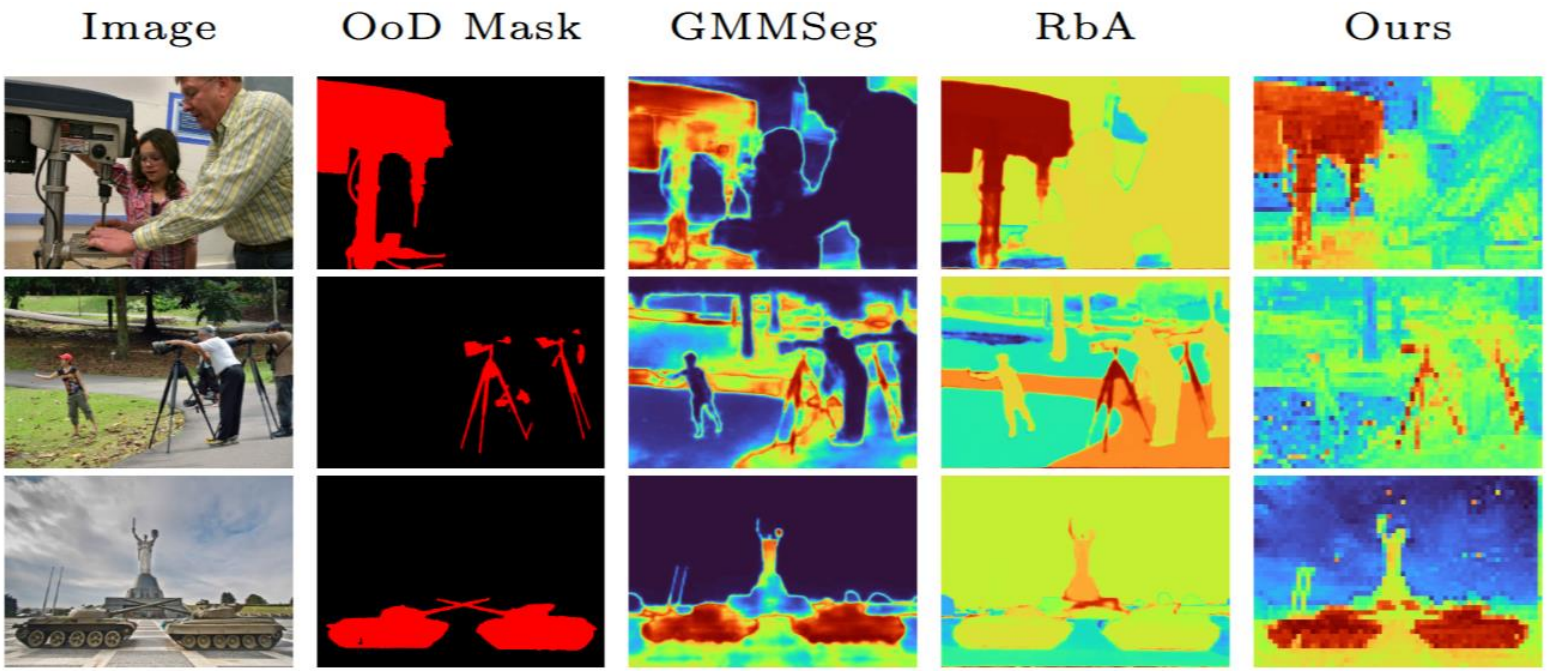
Method	AP	FPR
M2A	18.23	86.53
Maskomaly	28.23	82.68
GMMSeg	47.6	43.5
cDNP	62.35	39.20
RbA	66.82	82.42
Ours	63.03	36.50

< Quantitative comparisons on the ADE-OoD Dataset>

Experiments

- Qualitative Results

  타 모델보다 더 OoD Mask(GT)에 알맞은 OoD detecting을 하는 것을 알 수 있음



< Qualitative comparisons on the ADE-OoD Dataset >

감사합니다