arXiv, 2023

Generalized Category Discovery in Semantic Segmentation

Zhengyuan Peng. Qijian Tian. Jianqing Xu. Yizhang Jin. Xuequan Lu. Xin Tan. Yuan Xie. Lizhuang Ma.

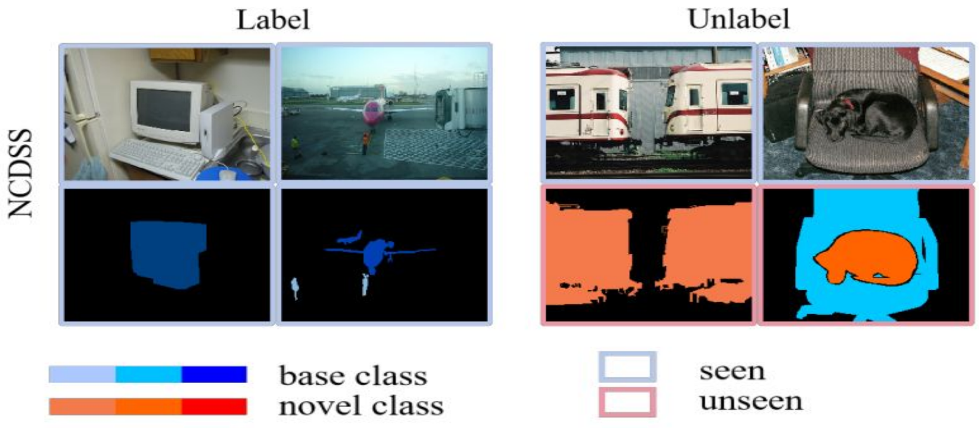
- Problem/Objective
  - Segment unlabeled images
- Contribution/Key Idea
  - GCDSS

전유진

- **NCDSS (Novel Class Discovery in Semantic Segmentation)**

-Goal.
discover novel classes based on prior knowledge from base classes.

-Problem.
each unlabeled image in the unlabeled set must contain at least one novel class.



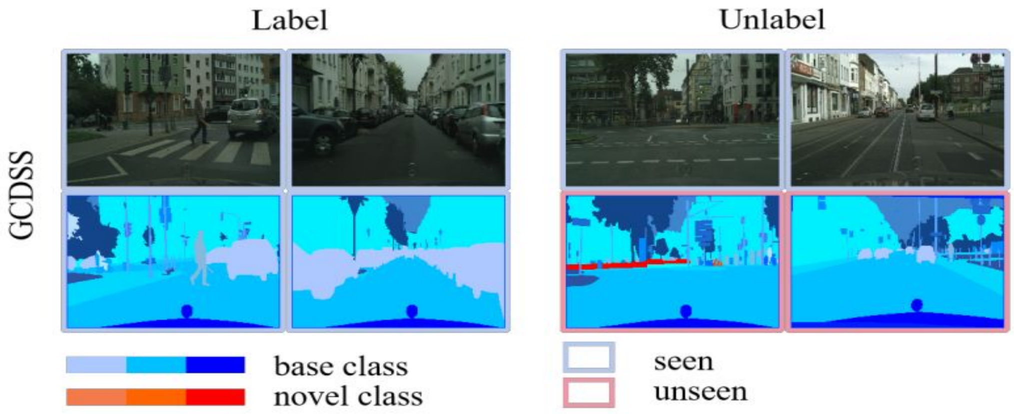
전유진

[Generalized Category Discovery in Semantic Segmentation]

- **GCDSS (General Category Discovery in Semantic Segmentation)**

-Goal.
discover novel classes by leveraging labeled data of base classes and unlabeled data.

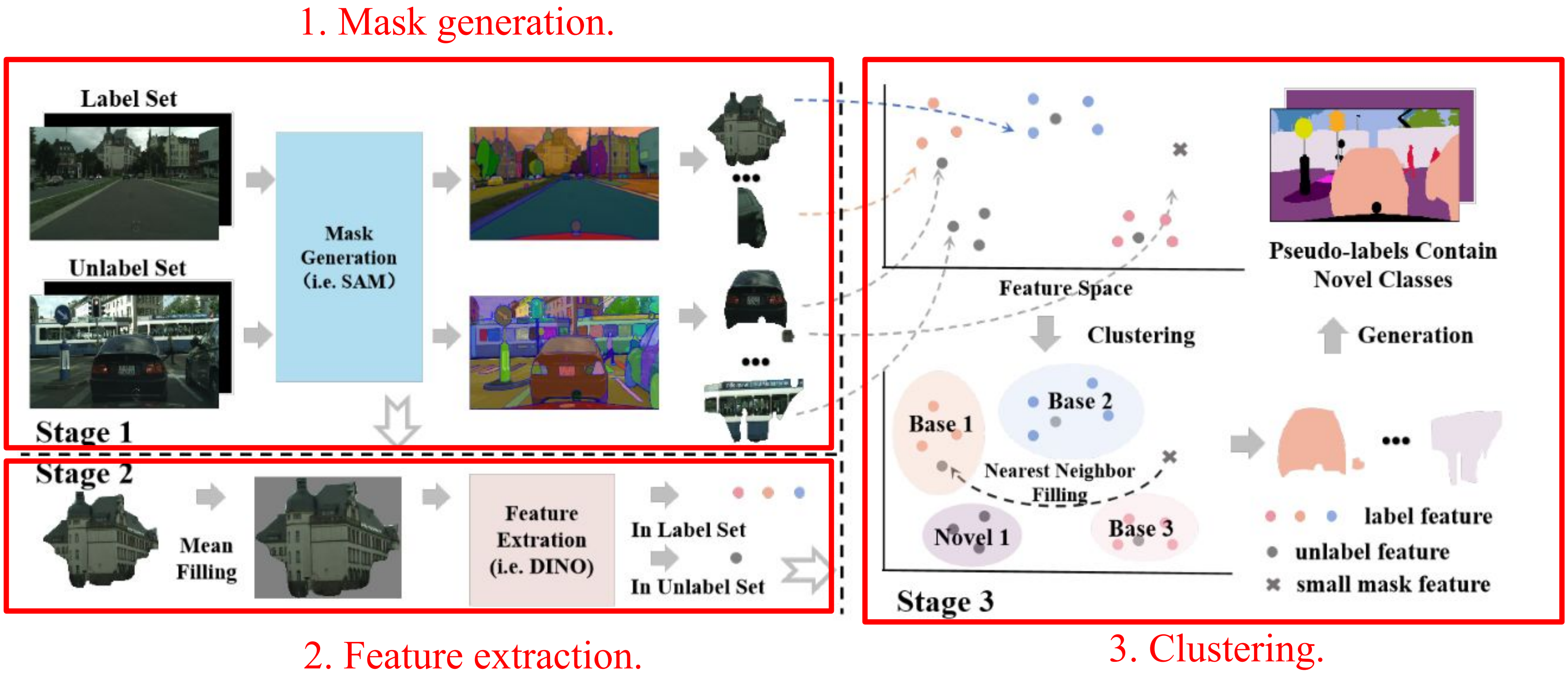
-Feature.
1. There is no prerequisite for prior knowledge mandating the existence of at least one novel class in each unlabeled image.
2. GCDSS broadens the segmentation scope beyond foreground objects.



전유진

[Generalized Category Discovery in Semantic Segmentation]

- **GCDSS**

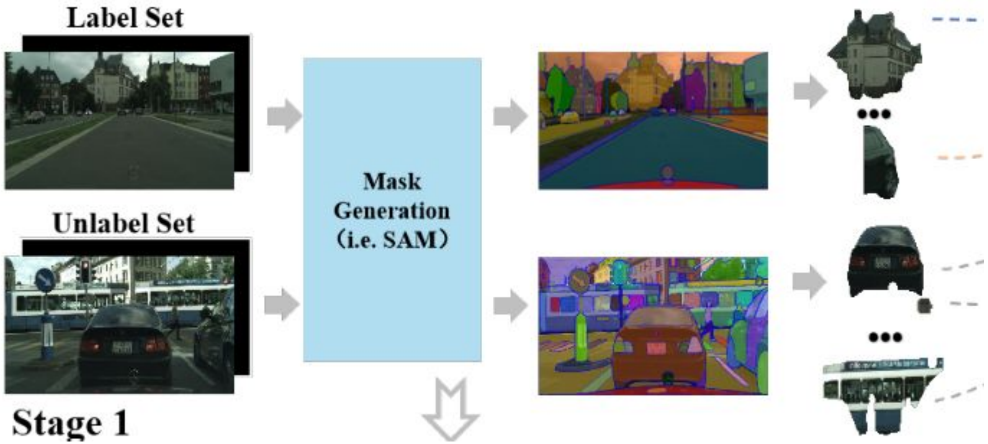


전유진

- **Mask generation**

-Input image. I

-Non-overlapping Masks. $M = \{m_1, m_2, ..., m_n\}$, $m_i \cap m_j = \emptyset$ for all $i \neq j$



($I = \sum m_i$ 되도록 SAM 보완.)

- **Feature extraction**

-Input image. I

-Non-overlapping Masks. $M = \{m_1, m_2, \ldots, m_n\}$, $m_i \cap m_j = \emptyset$ for all $i \neq j$

$$\Big\downarrow f(\cdot)$$

-Features. $F = \{f_1, f_2, \ldots, f_n\}$

전유진

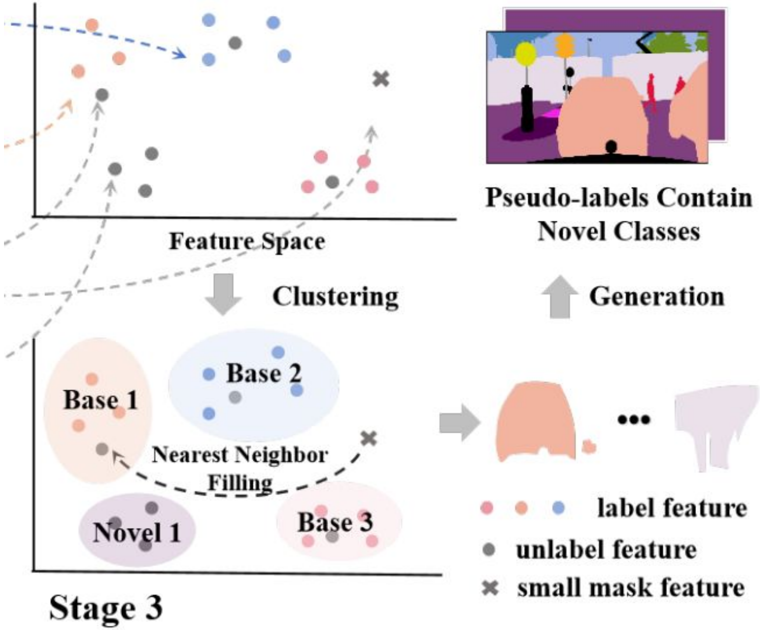
- **Clustering**

-Input image. I

-Non-overlapping Masks. $M = \{m_1, m_2, ..., m_n\}$, $m_i \cap m_j = \emptyset$ for all $i \neq j$

-Features. $F = \{f_1, f_2, ..., f_n\}$

-Labels. $L = \{l_1, l_2, ..., l_n\}$



$\therefore$ Final segmentation map = $\sum_{i=1}^{n} m_i \times l_i$

전유진

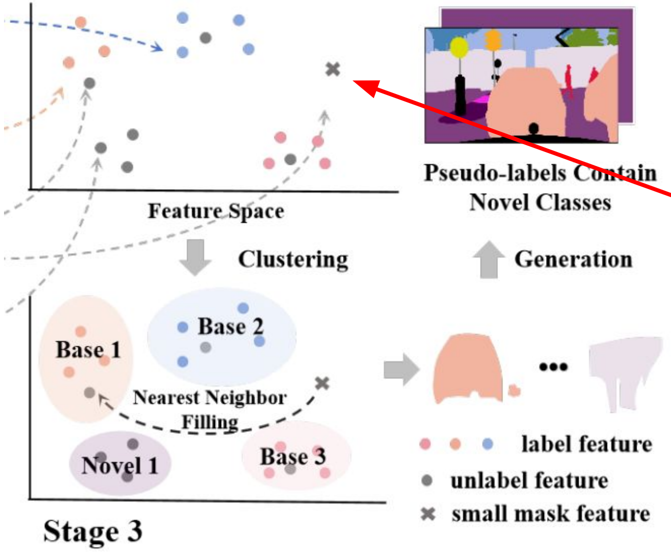
[Generalized Category Discovery in Semantic Segmentation]

- **Problem**

Hard to classify small masks that lack distinct features.

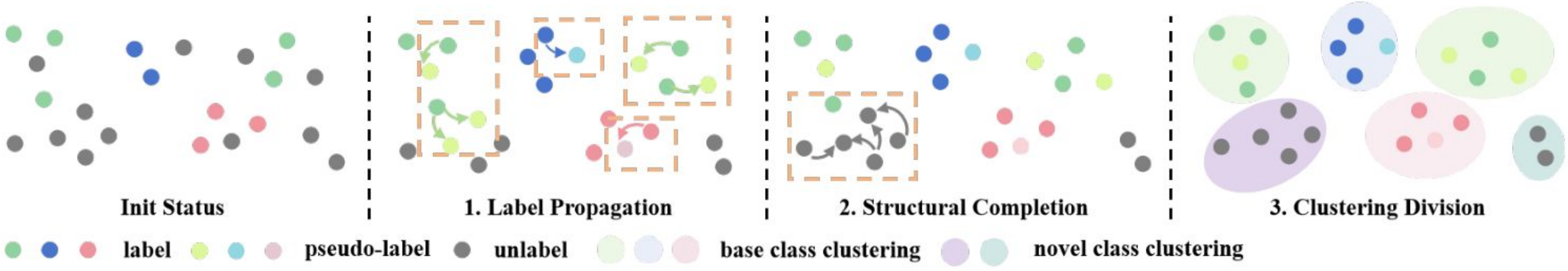
- **Solution**

Neighborhood Relations-Guided Mask Clustering Algorithm (NeRG-MaskCA).



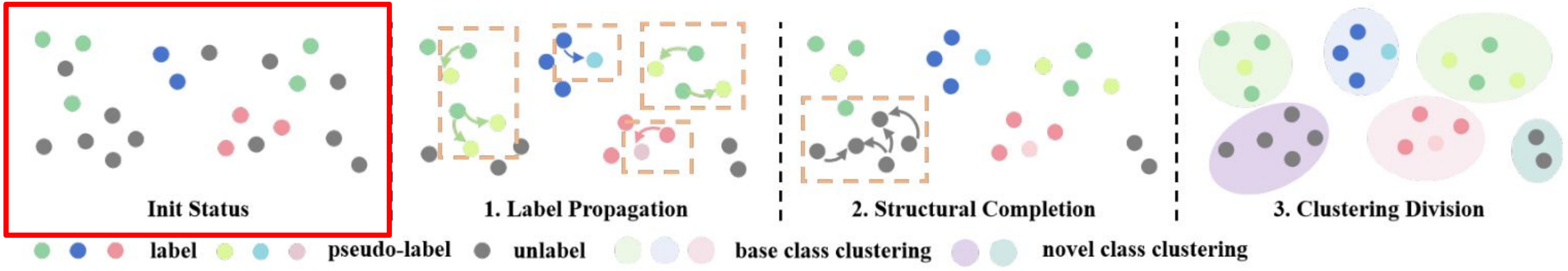
문제

- **NeRG-MaskCA**


Init Status
1. Label Propagation
2. Structural Completion
3. Clustering Division
label
pseudo-label
unlabel
base class clustering
novel class clustering

[Generalized Category Discovery in Semantic Segmentation]

- **Init status**



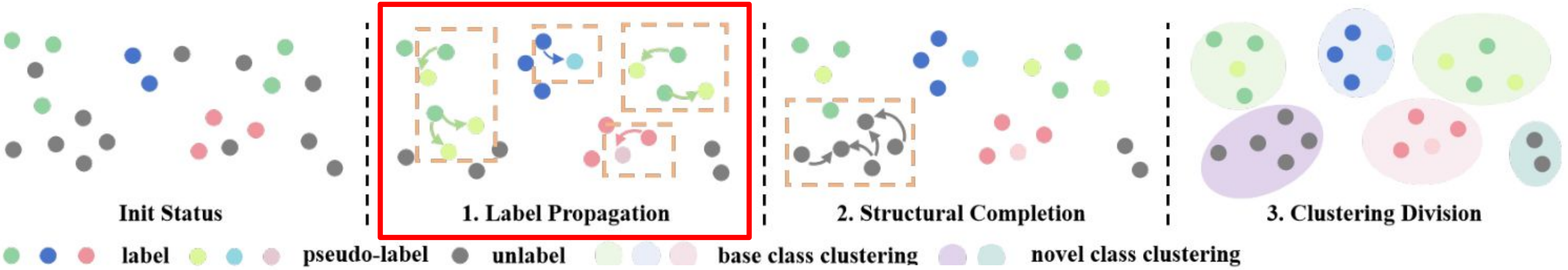
Algorithm 1 NeRG-MaskCA

Input: M_u, M_l, F,W,$L(M_l)$, where $M_u \cup M_l = M$
Output: $L(M_u)$
$p(m_u) \leftarrow 0$ for $m_u \in M_u$, $p(m_l) \leftarrow 1$ for $m_l \in M_l$ ▷ **Init**
for $x_u \sim M_u$ **do**
    for $m' \in M_u \cup M_l$ **do**
        dis$(m_u, m') \leftarrow ||F(m_u) - F(m')||_2$
    end for
    find and save top-k nearest mask of x_u
end for

전유진

[Generalized Category Discovery in Semantic Segmentation]

- **Label propagation**



-pseudo-label formula.

$$l = \begin{cases} \operatorname*{argmax}_{c} \left(\sum_{i=1}^{k} p_i \cdot \mathbb{1}_{\{label_i = c\}} \right) \\ \qquad , \text{if } \max \left(\sum_{i=1}^{k} p_i \cdot \mathbb{1}_{\{label_i = c\}} \right) > \theta \\ unlabel, \text{otherwise}, \end{cases}$$

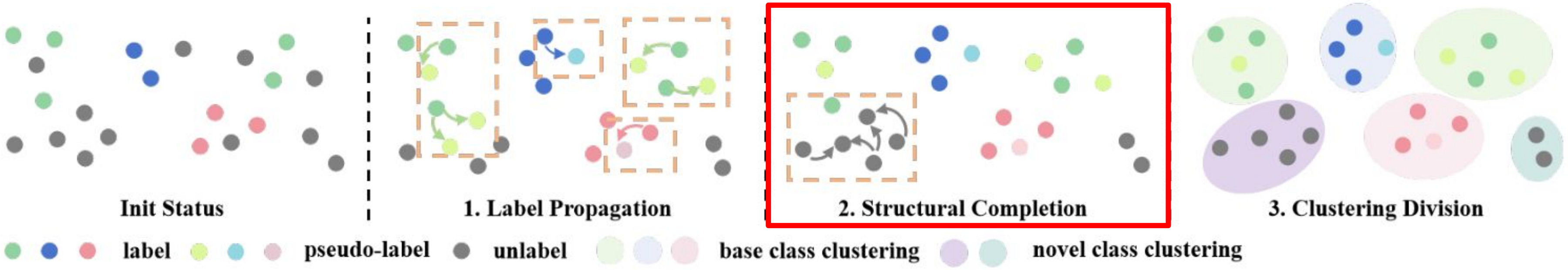
-update formula.

$$p = \begin{cases} \left(\sum_{i=1}^{k} p_i \cdot \mathbb{1}_{\{label_i = c\}} \right) \\ \qquad , \text{if } \max \left(\sum_{i=1}^{k} p_i \cdot \mathbb{1}_{\{label_i = c\}} \right) > \theta \\ 0, \text{otherwise}. \end{cases}$$

전유진

[Generalized Category Discovery in Semantic Segmentation]

- **Structural Completion**

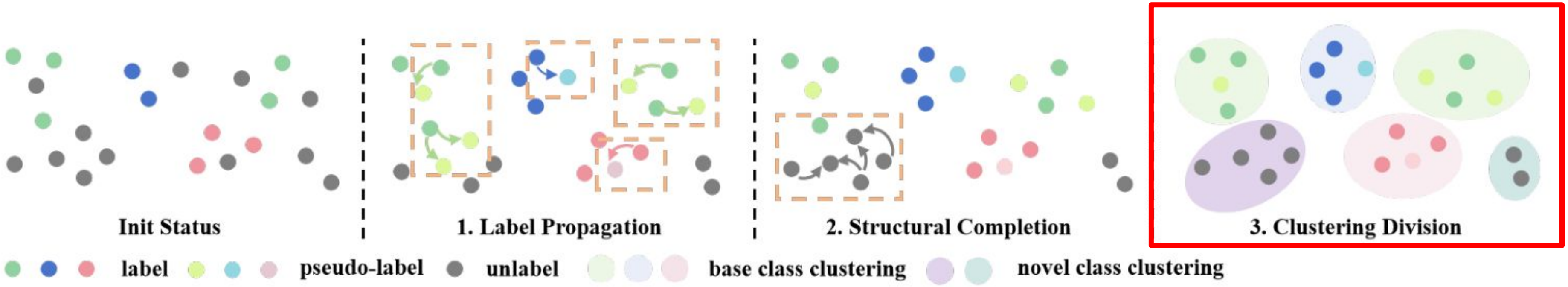


-elimination formula.

$$l = unlabel, \text{if} \left(\sum_{i=1}^{k} \cdot \mathbb{1}_{\{label_i = unlabel\}} \right) > \theta$$

전유진

- **Clustering Divison**



전유진

- **Experiment**

Labeled set D_l contains only the base classes.

Unlabeled set D_u includes both the base classes and novel classes.

Comb.	Novel Classes	Num / Pixel Area in Unlabel Set
1	Rider, Truck, Bus, Train	1816 / 1.31%
2	Rider, Bus, Train, Motor.	1805 / 1.05%
3	Wall, Truck, Bus, Train	1767 / 2.08%
4	Wall, Bus, Train, Motor.	1876 / 1.82%
5	Fence, Truck, Bus, Train	1986 / 2.38%

Table 1. **Cityscapes-GCD.** Our dataset includes five combinations, each with a labeled set (1390 images) and an unlabeled set (2085 images). It features 15 base classes and 4 novel classes. We also provide detailed information on the novel classes in the unlabeled set, including image number (Num) and pixel area proportion (Pixel Area).

Combination	Baseline			NeRG-MaskCA		
	Base Class	Novel Class	Avg Class	Base Class	Novel Class	Avg Class
Comb. 1	31.99	3.38	25.97	46.12	30.61	42.86
Comb. 2	28.38	2.36	22.9	46.62	28.94	42.90
Comb. 3	31.01	2.10	24.92	46.42	30.74	43.12
Comb. 4	32.3	3.86	26.31	46.84	28.03	42.88
Comb. 5	28.91	5.88	24.06	45.65	33.18	43.02
Average mIoU	30.52	3.52	24.83	**46.33**	**30.30**	**42.96**

Table 2. **Comparison of the baseline and NeRG-MaskCA across five class combinations.** NeRG-MaskCA outperforms the baseline compared to the five class combinations.

- **Ablation study**

Clustering Div.	Label Prop.	Struct Comp.	mIoU (%) (Base / Novel / Avg)
✓	-	-	30.52 / 3.52 / 24.83
✓	✓	-	46.31 / 23.92 / 41.60
✓	✓	✓	**46.33 / 30.30 / 42.96**

Table 3. **Comparison of components.**

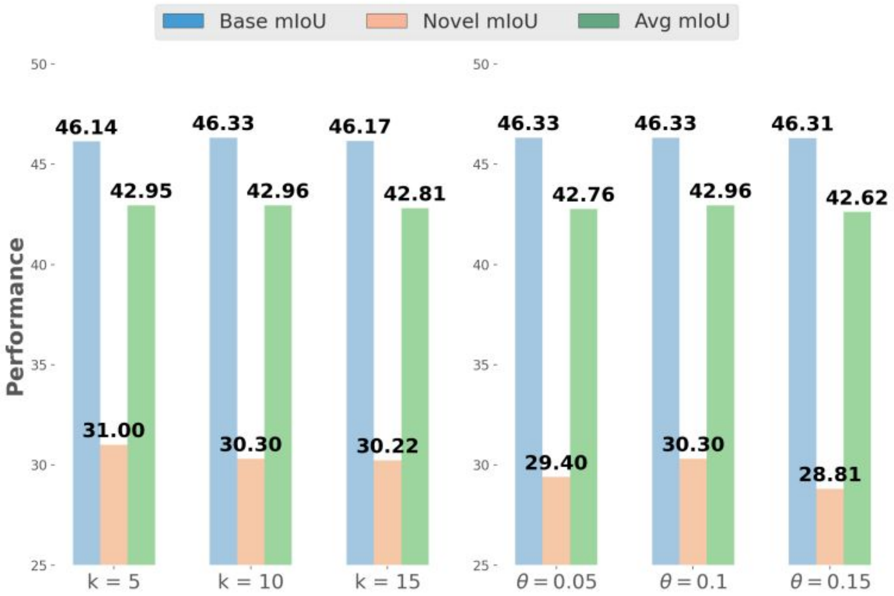


Figure 5. **Parameter analysis of k and θ.** The nearest mask number k varies among 5, 10, and 15. The lower bound confidence θ for pseudo-label changes among 0.05, 0.10, and 0.15. The performance of our approach is relatively stable.

전유진