# A²LC: Active and Automated Label Correction for Semantic Segmentation

Youjin Jeon*, Kyusik Cho*, Suhan Woo, Euntai Kim†
Yonsei University

## Introduction

- Task: **Label acquisition for semantic segmentation**
  - Challenge: High cost and labor-intensive nature of pixel-wise annotation
  - Promising solution: Active Label Correction with foundation models (ALC)

- Two key challenges of **ALC**
  1. Redundant queries for similar patterns
  2. Class imbalance challenge

- Main contributions of **A²LC**
  1. Label Correction Module (**LCM**) automatically corrects noisy labels exhibiting similar features with human-corrected masks.
  2. Adaptively Balanced Confidence in label (**ABC**) increases the sampling frequency of tail classes.

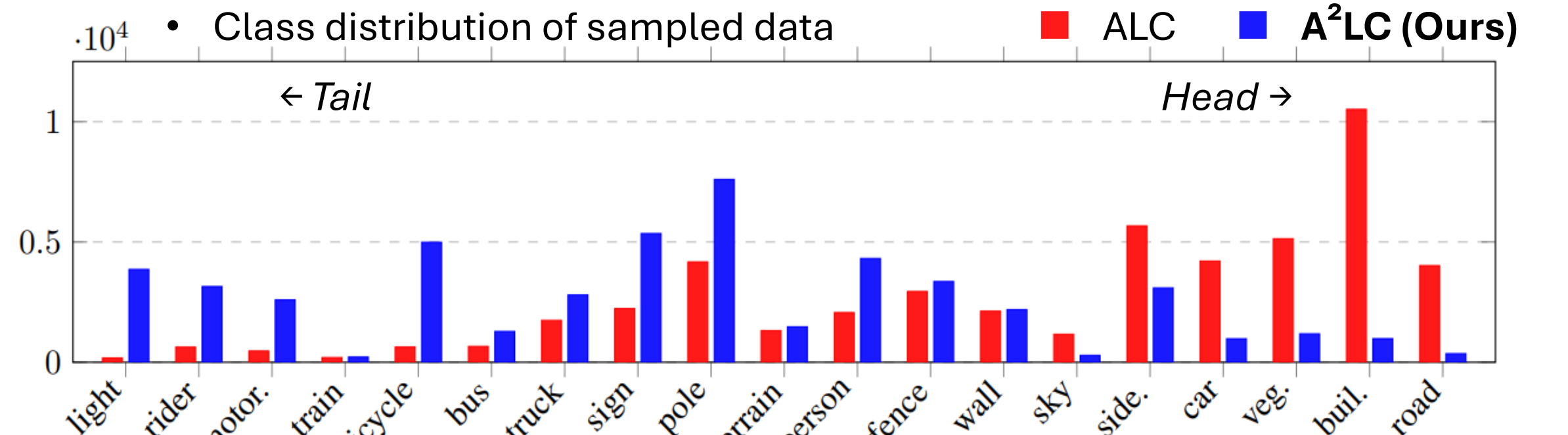## Preliminaries: ALC Framework

### Overview of ALC framework

1. Initial pseudo-labels are generated by the foundation model.
2. In each round, top-$B$ informative pixels are sampled by acquisition function using deep learning model and queried to annotator.
3. After manual corrections, both the pseudo-labels and model are updated, completing a single round of the correction cycle.

### Two key challenges of **ALC**

1. Redundant queries lead to inefficient resource utilization.



■ Traffic sign  ■ Traffic light
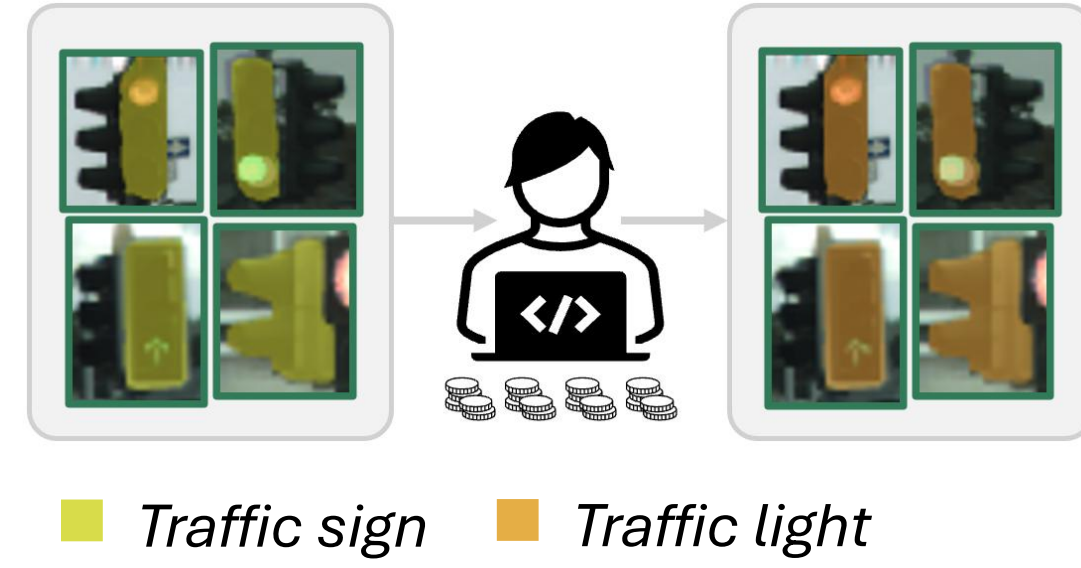
2. Sampling bias toward head classes results in suboptimal corrections.

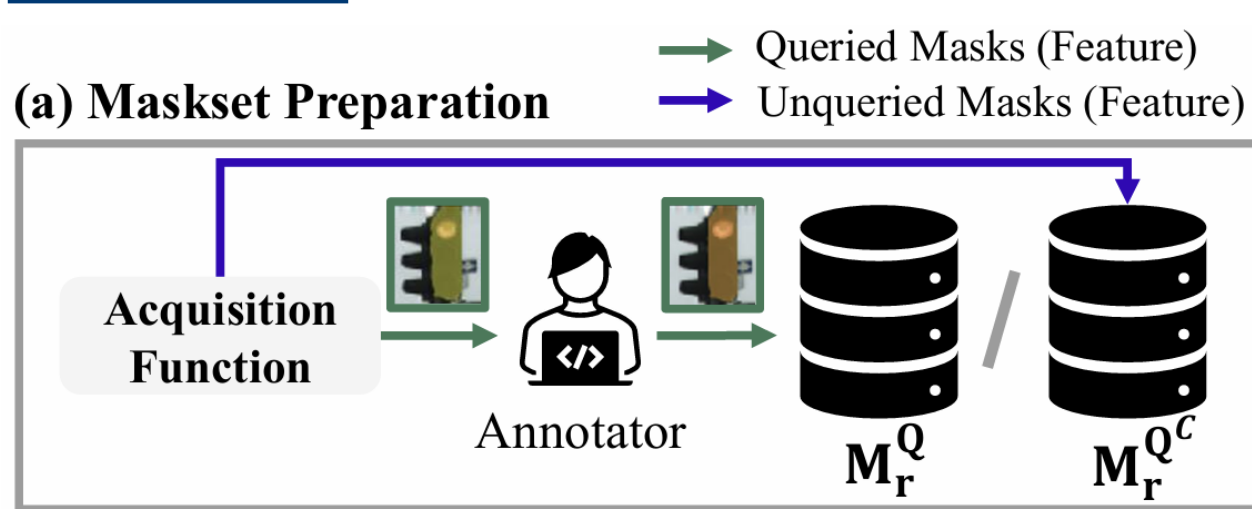- Class distribution of sampled data   ■ ALC   ■ A²LC (Ours)



## A²LC Framework

### Overview of A²LC framework

- A²LC framework is a semi-automated label correction framework for semantic segmentation built upon cascading stages of manual and automatic correction.
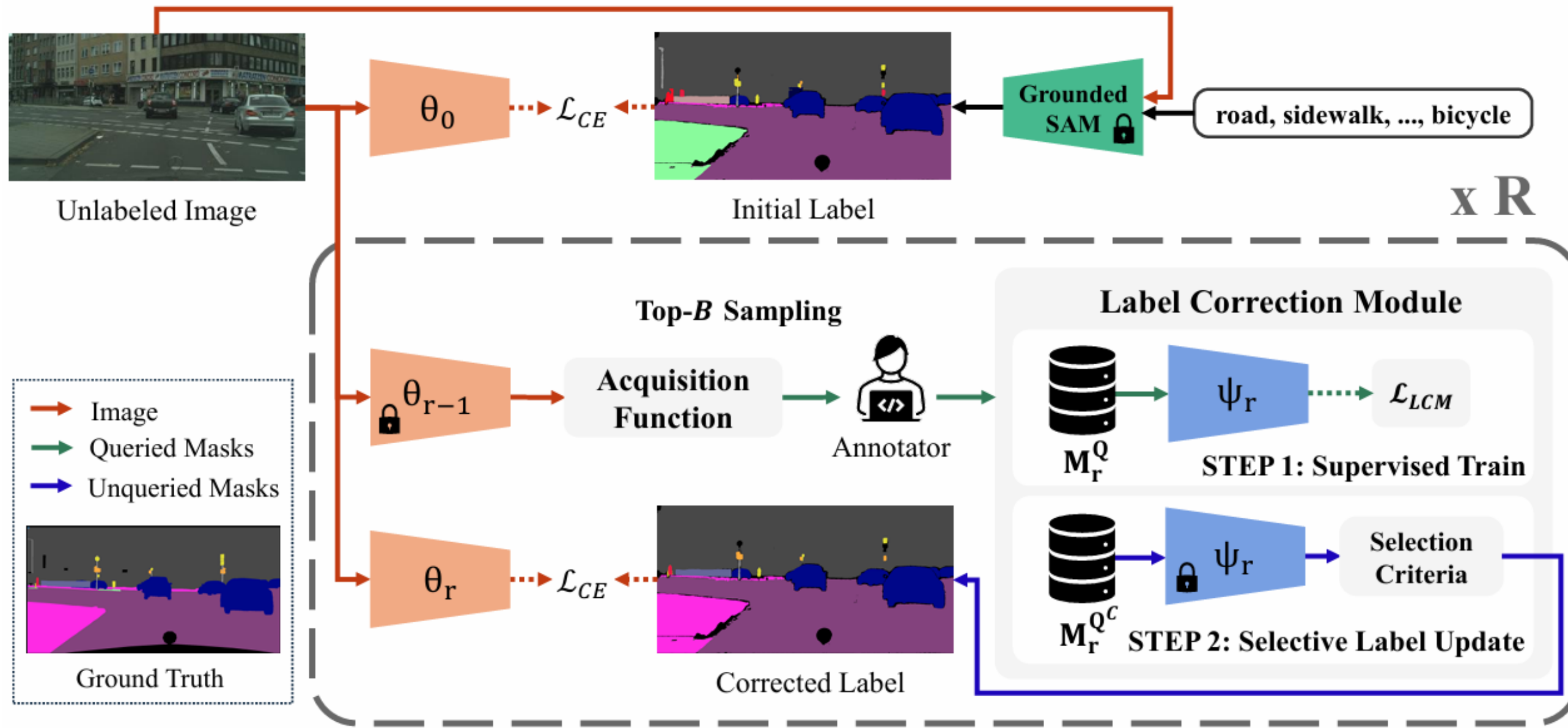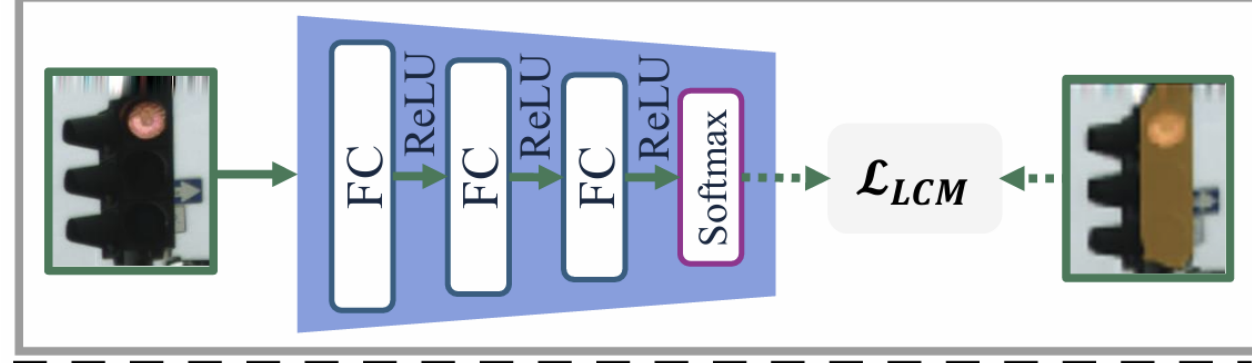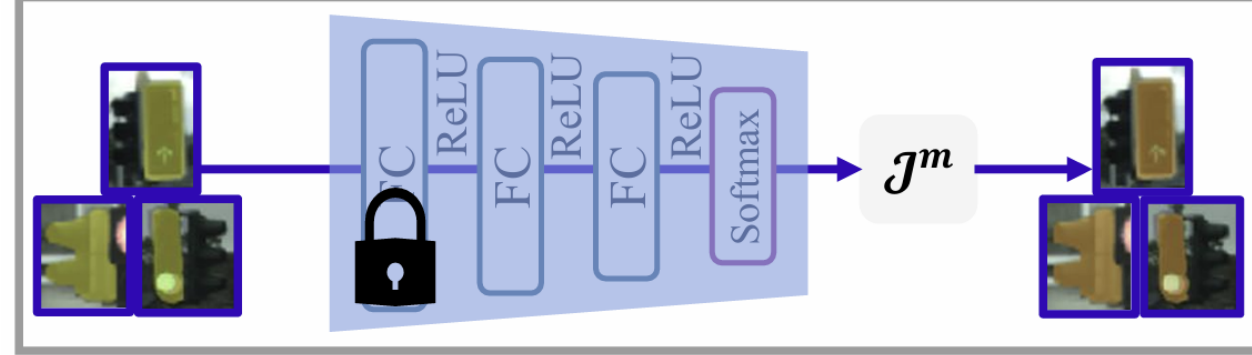


### Main contributions of A²LC

#### 1. LCM



(a) Maskset Preparation
(b) STEP 1: Supervised Train
(c) STEP 2: Selective Label Update

- LCM performs automatic correction by propagating human-provided labels beyond the queried samples.

- LCM correction is applied only to the masks that satisfy all selection criteria.

- **Selection criteria** are designed to identify predictions with high reliability, so that automatic label updates are applied exclusively to the subset of confidently inferred labels.

$$\mathcal{J}^m = \mathcal{J}_1^m \wedge \mathcal{J}_2^m \wedge \mathcal{J}_3^m$$
$$\mathcal{J}_1^m = \mathbb{I}\left(\max_{c \in \mathcal{C}} \psi_r(c; m) \geq \tau\right)$$
$$\mathcal{J}_2^m = \mathbb{I}\left(\hat{y}_{\psi_r}(m) \notin \{c \mid \mathrm{rank}(c) \geq (1-\alpha) \cdot |\mathcal{C}|, \ c \in \mathcal{C}\}\right)$$
$$\mathcal{J}_3^m = \mathbb{I}\left(\hat{y}(m) \neq \arg\max_{c \in \mathcal{C}} \mathrm{rank}(c)\right)$$

#### 2. ABC

- ABC guides both correction stages toward improved class balance by incorporating pixel-wise adaptive class weight into the acquisition function.

$$a_{\mathrm{ABC}}(m; \theta) := \sum_{x \in m} \frac{f_\theta(x) \cdot f_\theta(m')}{\|f_\theta(x)\| \|f_\theta(m')\|} \cdot a_{\mathrm{ABC}}(x; \theta) \qquad a_{\mathrm{ABC}}(x; \theta) := w(x) \cdot a_{\mathrm{CIL}}(x; \theta)$$

- **Adaptive class weight** is composed of two components, class rarity score and dataset imbalance score.
  - **Class rarity score** prioritizes pixels belonging to tail classes during sampling.
  - **Dataset imbalance score** adaptively emphasizes the class rarity score according to the pseudo-label statistics at each round.

$$w(x) := \hat{w}(x)^{\mathrm{KL}^3(\mathbb{P}_{\mathrm{dist}} \| \mathbb{U}_{\mathrm{dist}})} \qquad \hat{w}(x) := \frac{\min_{c \in \mathcal{C}} |\{x' \in \mathcal{M} : \hat{y}(x') = c\}|}{|\{x' \in \mathcal{M} : \hat{y}(x') = \hat{y}(x)\}|} \qquad \mathrm{KL}(\mathbb{P}_{\mathrm{dist}} \| \mathbb{U}_{\mathrm{dist}}) = \sum_{c \in \mathcal{C}} \mathbb{P}(c) \log \frac{\mathbb{P}(c)}{\mathbb{U}(c)}$$

## Experimental Results

### Quantitative results

- **High efficiency**: Requires only 20% (Cityscapes) and 60% (PASCAL) of the budget.
- **Strong effectiveness**: Achieves +27.23% (Cityscapes) and +14.30% (PASCAL) performance gains under the same budget.
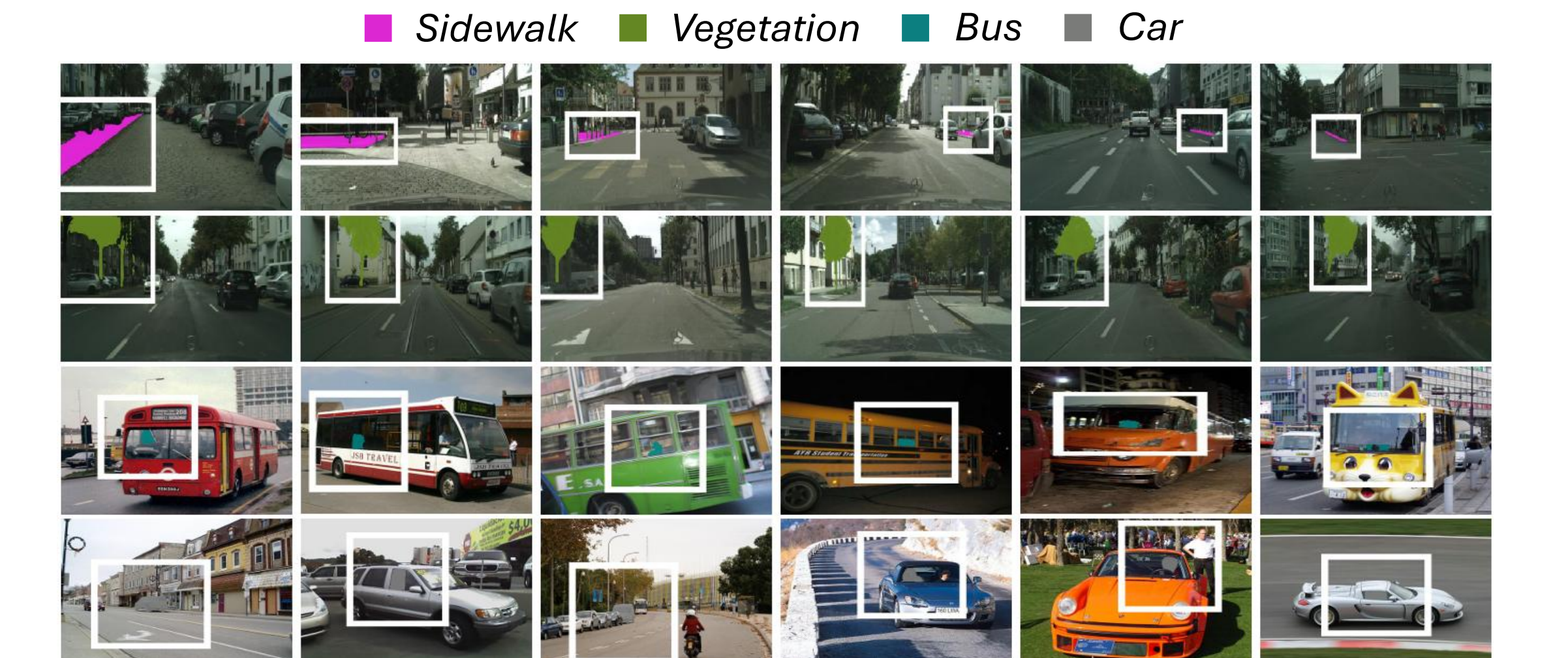
| Dataset | mIoU (%) | Methods | Init. | 1R | 2R | 3R | 4R | 5R |
|---|---|---|---|---|---|---|---|---|
| Cityscapes | Data | ALC | 50.68±0.00 | 62.28±1.25 | 65.36±0.56 | 66.19±0.60 | 66.87±0.29 | 67.01±0.30 |
| | | A²LC (Ours) | 50.68±0.00 | 70.01±1.05 | 75.60±0.58 | 80.84±0.53 | 82.69±0.92 | 85.26±0.55 |
| | Model | ALC | 51.55±0.71 | 56.61±0.42 | 58.27±0.63 | 58.58±0.08 | 58.52±0.26 | 58.59±0.05 |
| | | A²LC (Ours) | 51.55±0.71 | 60.83±0.66 | 63.89±0.44 | 67.50±0.31 | 68.87±0.78 | 70.51±0.32 |
| PASCAL | Data | ALC | 58.63±0.00 | 68.19±0.34 | 72.72±0.16 | 74.84±0.29 | 76.41±0.70 | 77.06±0.73 |
| | | A²LC (Ours) | 58.63±0.00 | 67.49±2.03 | 74.88±1.45 | 80.88±0.83 | 84.81±0.48 | 88.08±0.44 |
| | Model | ALC | 56.94±0.44 | 62.11±0.61 | 64.12±0.31 | 64.15±0.68 | 65.00±0.23 | 65.48±0.83 |
| | | A²LC (Ours) | 56.94±0.44 | 60.87±2.93 | 64.08±2.43 | 66.45±0.95 | 67.76±0.06 | 68.42±0.87 |

### Qualitative results

- Constructed pseudo-labels



Unlabeled Image   Init.   ALC   **A²LC (Ours)**   Ground Truth
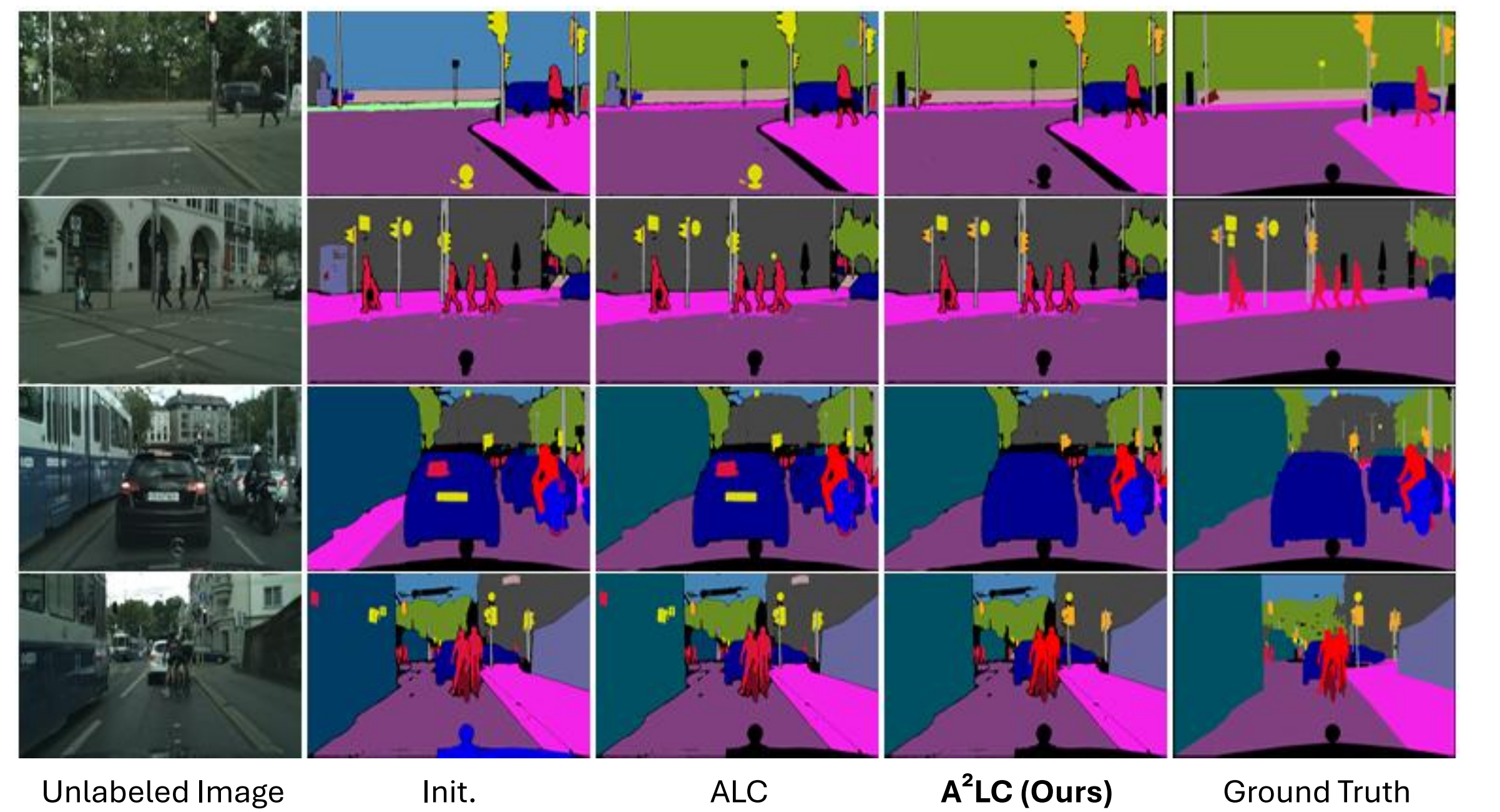
- Visualization of LCM-corrected masks

■ *Sidewalk*  ■ *Vegetation*  ■ *Bus*  ■ *Car*



### Ablation study

| Methods | | | Data mIoU (%) | | | | | Model mIoU (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LCM | ABC | Mask | 1R | 2R | 3R | 4R | 5R | 1R | 2R | 3R | 4R | 5R |
| | | | 62.60 | 65.20 | 66.03 | 66.85 | 66.86 | 56.86 | 58.93 | 58.63 | 58.59 | 58.59 |
| ✓ | | | 65.66 | 71.53 | 76.00 | 78.38 | 80.59 | 59.02 | 61.39 | 64.19 | 65.58 | 66.94 |
| ✓ | ✓ | | 67.31 | 70.06 | 74.49 | 78.32 | 81.58 | 59.09 | 60.86 | 62.05 | 65.27 | 67.75 |
| ✓ | ✓ | | 70.62 | 77.29 | 80.39 | 82.80 | 84.15 | 60.32 | 65.09 | 67.36 | 68.71 | 70.11 |
| ✓ | ✓ | ✓ | 71.04 | 76.08 | 81.13 | 83.53 | 85.57 | 61.06 | 64.39 | 67.83 | 69.63 | 70.88 |