

A²LC: Active and Automated Label Correction for Semantic Segmentation

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

Yonsei University



YONSEI University
CI Computational
Intelligence Laboratory



AAAI-26 / IAAI-26 / EAAI-26
JANUARY 20-27, 2026 | SINGAPORE

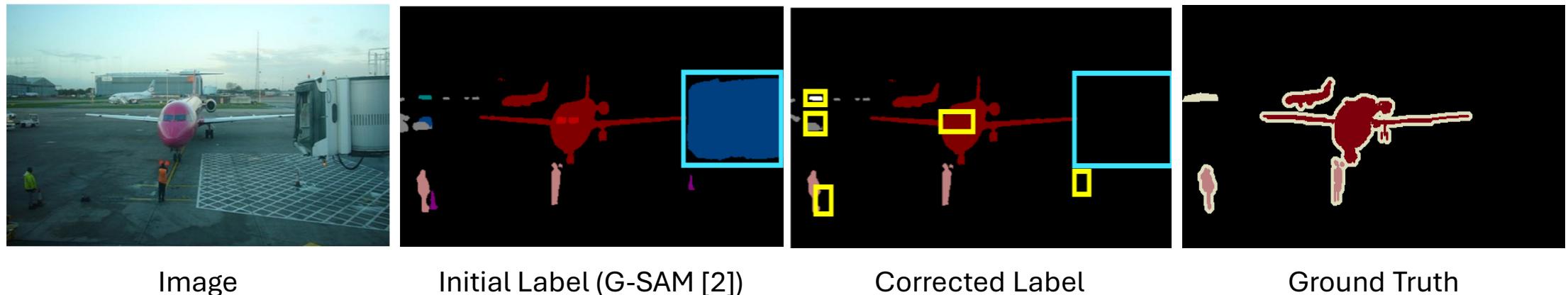
Preliminaries: Active Label Correction

Active Label Correction with foundation models (**ALC**) [1] has emerged as a promising solution to the aforementioned difficulty.

ALC is a method that

- (1) generates **initial pseudo-labels** using a foundation model,
- (2) **selects** noisy labels through a deep learning model and
- (2) **corrects** only those labels via human review in an iterative manner.

cf. Figures are from [1]



[1] KIM, Hoyoung, et al. Active label correction for semantic segmentation with foundation models. In Forty-first International Conference on Machine Learning, 2024.

[2] REN, Tianhe, et al. Grounded sam: Assembling open-world models for diverse visual tasks. arXiv preprint arXiv:2401.14159, 2024.

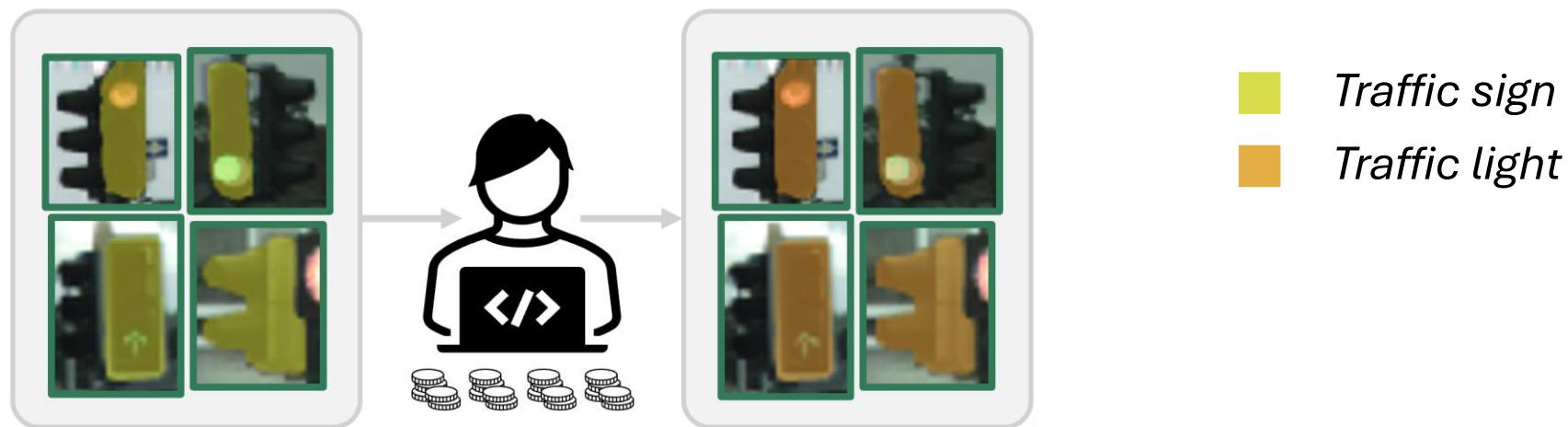
Two key challenges of ALC

The first challenge is

redundant queries for similar patterns, leading to inefficient resource utilization.

What are redundant queries?

Label confusion predominantly occurs among similar categories, that are naturally clustered within comparable acquisition score ranges and thus queried together.



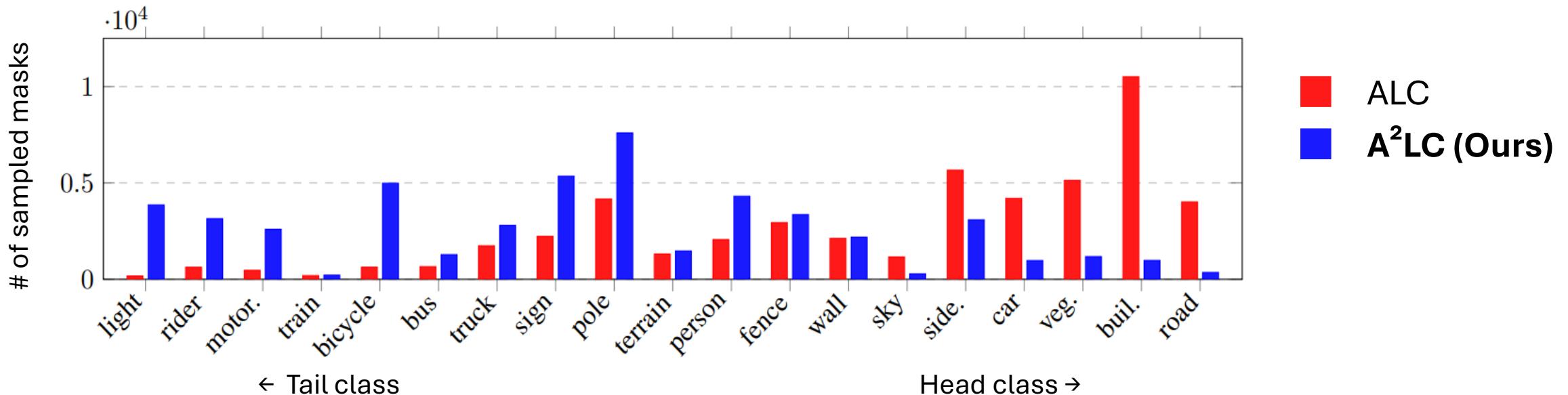
Two key challenges of ALC

The second challenge is

sampling bias toward head classes, resulting in suboptimal corrections.

How does sampling bias affect the results?

Corrections overwhelmingly target head class pixels, even when their performance is already saturated, leaving noisy tail class labels uncorrected despite multiple correction rounds.

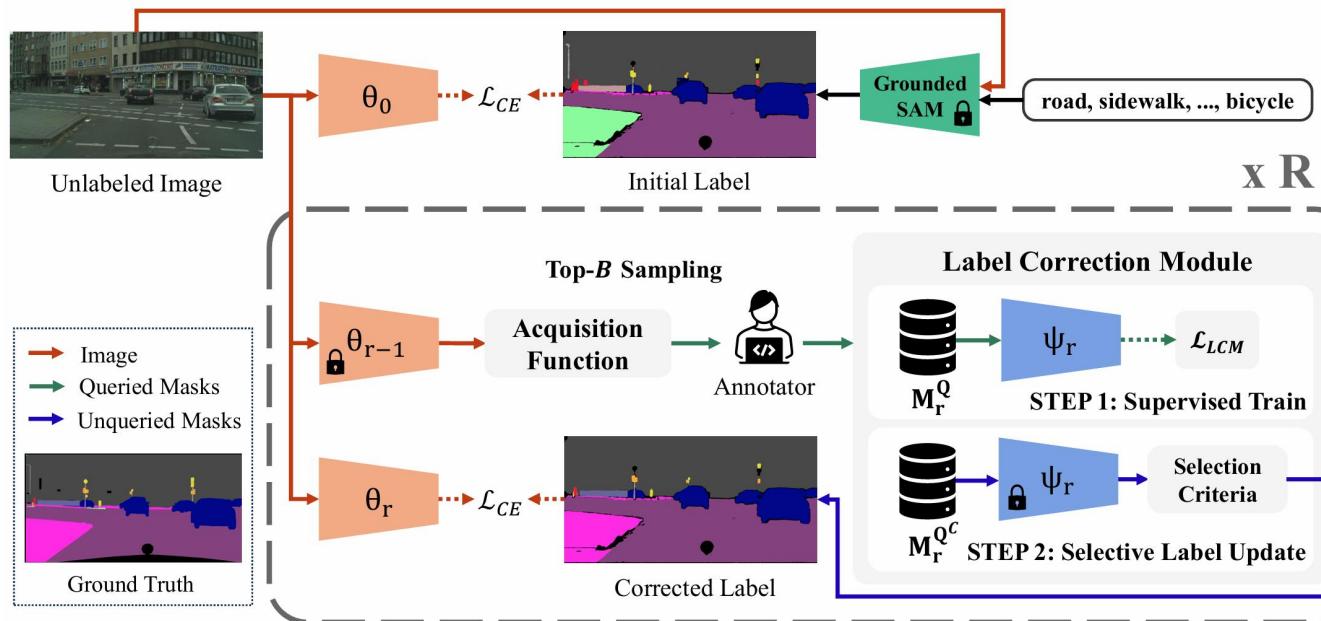


A^2LC Framework

We introduce an Active and Automated Label Correction framework (A^2LC), that effectively addresses the two aforementioned challenges.

A^2LC framework is

a **semi-automated label correction framework** for semantic segmentation, where manual and automatic correction stages operate in a cascaded manner.

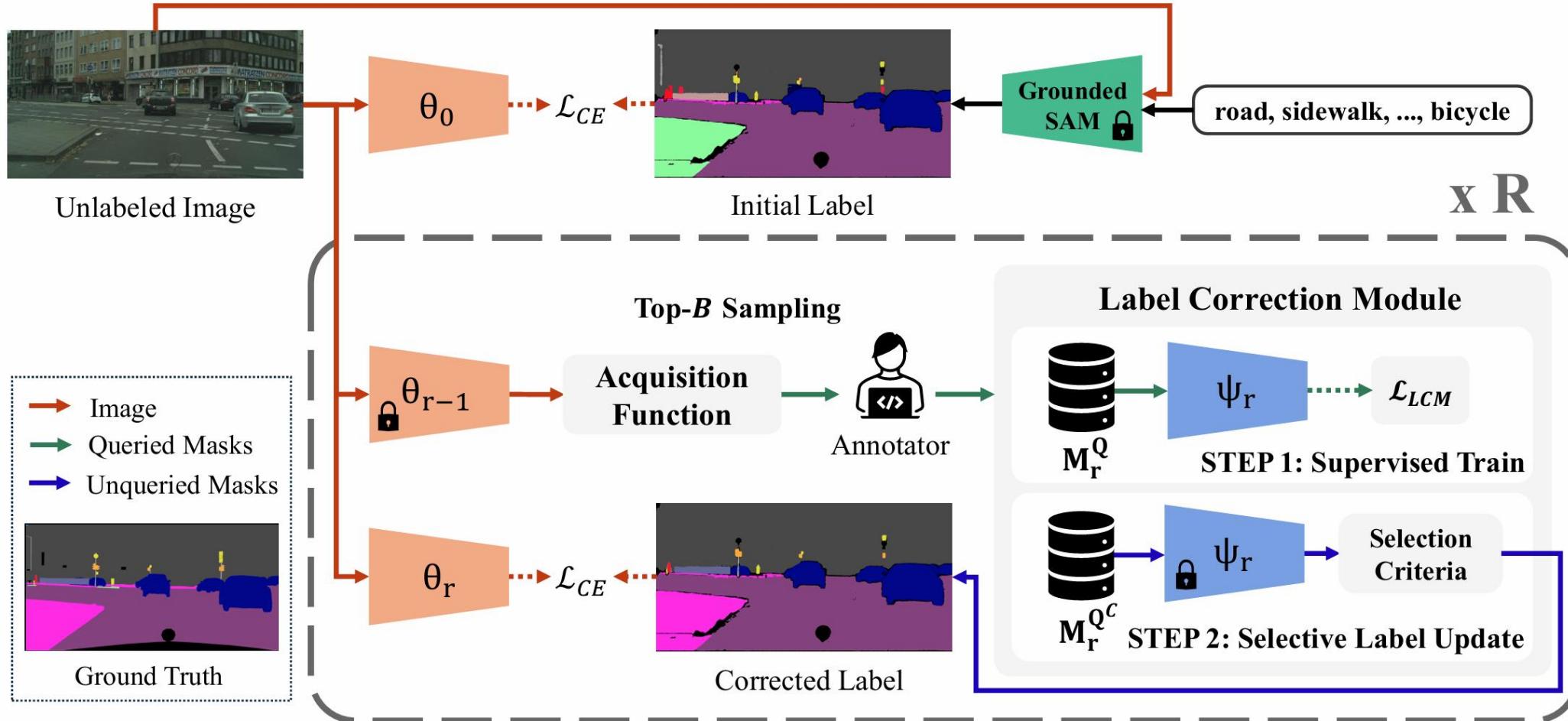


Algorithm 1: A^2LC Framework

Require: Budget size B , Final round R
Ensure: \mathcal{D}_R and θ_R

- 1: Prepare initial dataset \mathcal{D}_0 and maskset \mathcal{M}_0
- 2: Train model θ_0 with \mathcal{D}_0
- 3: **for** $r = 1$ to R **do**
- 4: **for** each $(m, \hat{y}(m)) \in \mathcal{M}_{r-1}$ **do**
- 5: Score all masks based on Eq. (12)
- 6: **end for**
- 7: Sample the top- B masks \mathcal{M}_r^Q based on Eq. (13)
- 8: $\mathcal{M}_r^Q \leftarrow \left\{ (m, y(m)) \mid m \in \mathcal{M}_r^Q, y(m) = \text{Oracle}(m) \right\}$
- 9: $\mathcal{M}_r^{Q^c} \leftarrow \mathcal{M}_{r-1} \setminus \mathcal{M}_r^Q$
- 10: $\mathcal{D}_r \leftarrow \text{LCM}(\mathcal{D}_{r-1}, \mathcal{M}_r^Q, \mathcal{M}_r^{Q^c})$
- 11: Train model θ_r with corrected \mathcal{D}_r
- 12: $\mathcal{M}_r \leftarrow \mathcal{M}_{r-1} \setminus \mathcal{M}_r^Q$
- 13: **end for**

Overview of A²LC framework



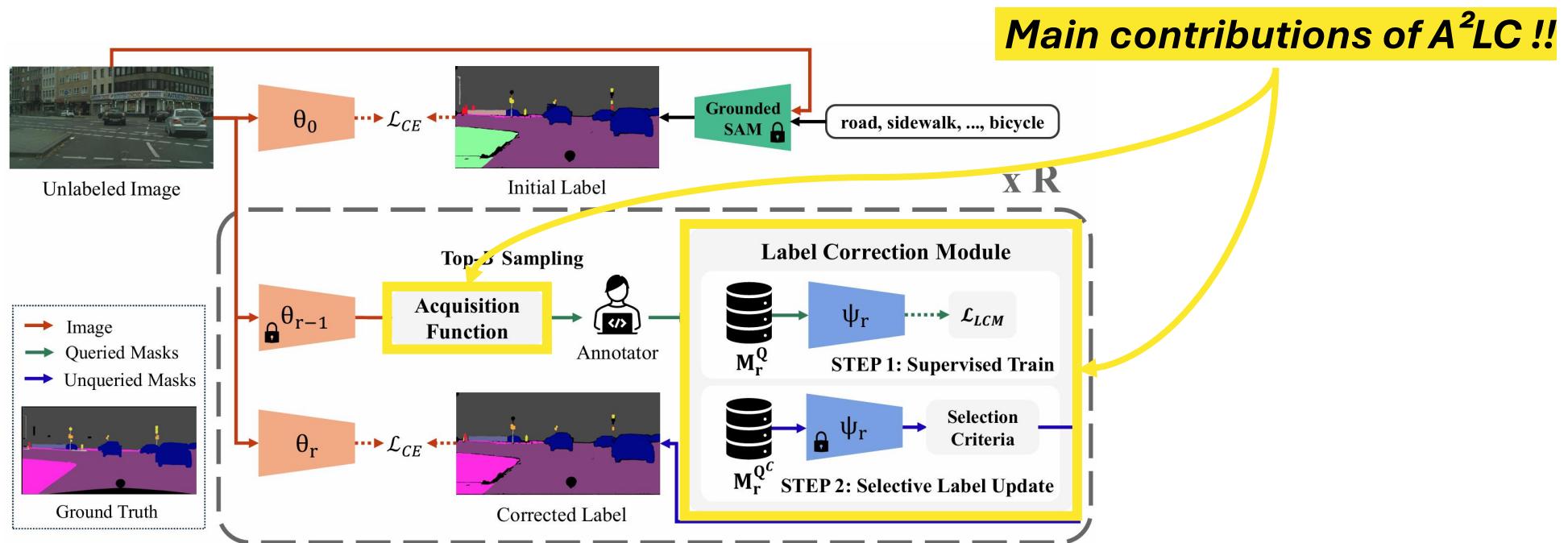
Main contributions of A²LC

The first contribution is

Label Correction Module (**LCM**) for additional automatic correction.

The second contribution is

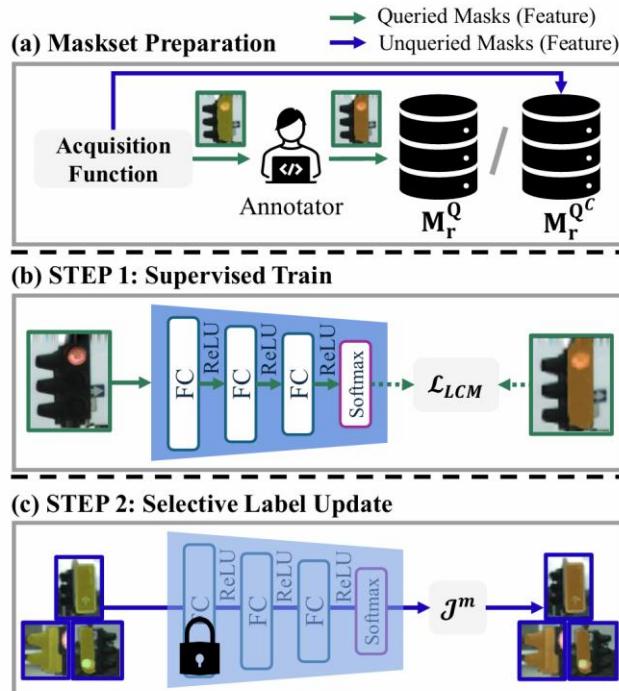
Adaptively Balanced Confidence in label (**ABC**) for class-balanced correction.



Label Correction Module (LCM)

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

LCM operates in two steps:
first, the model is **trained** using masks queried in the current round, and
second, it **corrects** masks that may be mislabeled but still remain unqueried.



Algorithm 2: Label Correction Module (LCM)

Require: Previous round dataset \mathcal{D}_{r-1} , Queried maskset \mathcal{M}_r^Q , Unqueried maskset $\mathcal{M}_r^{Q^c}$, Epochs E
Ensure: Updated dataset \mathcal{D}_r

- 1: \triangleright **Supervised train with queried masks**
- 2: **for** $e = 1, 2, \dots, E$ **do**
- 3: **for each** $(m, y(m)) \in \mathcal{M}_r^Q$ **do**
- 4: Train the model ψ_r based on Eq. (2)
- 5: **end for**
- 6: **end for**
- 7: \triangleright **Selective label update for unqueried masks**
- 8: **for each** $(m, \hat{y}(m)) \in \mathcal{M}_r^{Q^c}$ **do**
- 9: $\hat{y}_{\psi_r}(m) \leftarrow \arg \max_{c \in \mathcal{C}} \psi_r(c; m)$
- 10: **if** m satisfies *Selection_Criteria* (Eq. (3)) **then**
- 11: $\hat{y}(m) \leftarrow \hat{y}_{\psi_r}(m)$
- 12: **end if**
- 13: **end for**

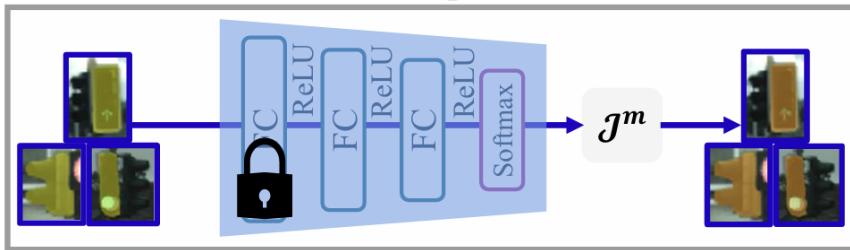
Label Correction Module (LCM)

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

What are selection criteria?

Selection criteria detect highly reliable predictions,
enabling automatic label updates exclusively for confidently inferred labels.

(c) STEP 2: Selective Label Update



```
7: ▷ Selective label update for unqueried masks
8: for each  $(m, \hat{y}(m)) \in \mathcal{M}_r^{Q^c}$  do
9:    $\hat{y}_{\psi_r}(m) \leftarrow \arg \max_{c \in \mathcal{C}} \psi_r(c; m)$ 
10:  if  $m$  satisfies Selection_Criteria (Eq. (3)) then
11:     $\hat{y}(m) \leftarrow \hat{y}_{\psi_r}(m)$ 
12:  end if
13: end for
```

$$\mathcal{J}^m = \mathcal{J}_1^m \wedge \mathcal{J}_2^m \wedge \mathcal{J}_3^m$$

$$\mathcal{J}_1^m = \mathbb{I}\left(\max_{c \in \mathcal{C}} \psi_r(c; m) \geq \tau\right)$$

$$\mathcal{J}_2^m = \mathbb{I}\left(\hat{y}_{\psi_r}(m) \notin \{c \mid \text{rank}(c) \geq (1 - \alpha) \cdot |\mathcal{C}|, c \in \mathcal{C}\}\right)$$

$$\mathcal{J}_3^m = \mathbb{I}\left(\hat{y}(m) \neq \arg \max_{c \in \mathcal{C}} \text{rank}(c)\right)$$

Adaptively Balanced Confidence in Label (ABC)

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

Adaptive class weight consists of two components:

- (1) **class rarity score**, that prioritizes tail class pixels during sampling, and
- (2) **dataset imbalance score**, that adjusts this weighting based on label statistics.

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

$$a_{\text{ABC}}(x; \theta) := w(x) \cdot a_{\text{CIL}}(x; \theta)$$

$$\begin{aligned} w(x) &:= \hat{w}(x)^{\text{KL}^3(\mathbb{P}_{\text{dist}} \parallel \mathbb{U}_{\text{dist}})} \\ \hat{w}(x) &:= \frac{\min_{c \in \mathcal{C}} |\{x' \in \mathcal{M} : \hat{y}(x') = c\}|}{|\{x' \in \mathcal{M} : \hat{y}(x') = \hat{y}(x)\}|} \end{aligned}$$

$$\text{KL}(\mathbb{P}_{\text{dist}} \parallel \mathbb{U}_{\text{dist}}) = \sum_{c \in C} \mathbb{P}(c) \log \frac{\mathbb{P}(c)}{\mathbb{U}(c)}$$

Experiments: Quantitative results

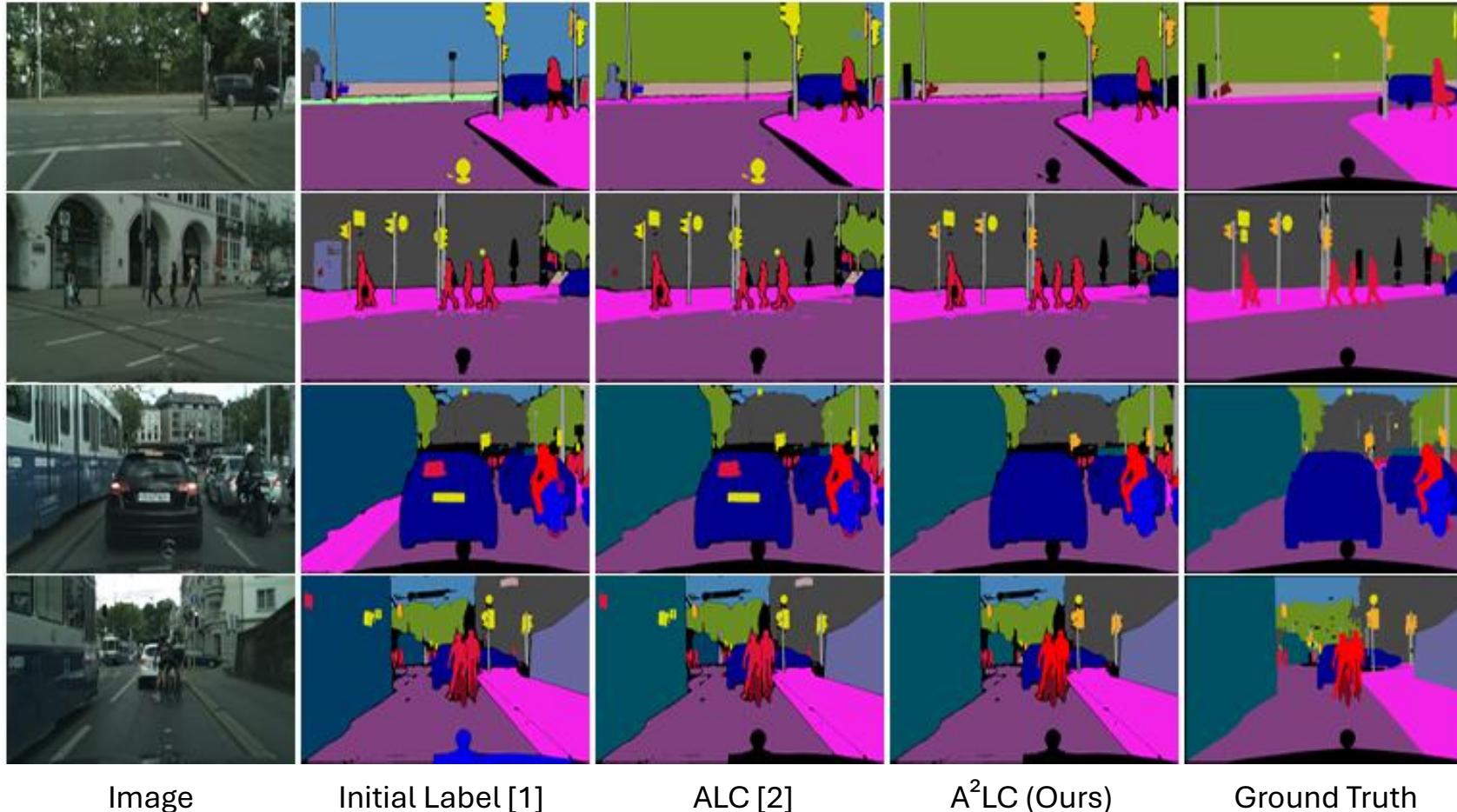
- **High efficiency:** Requires only 20% (Cityscapes) and 60% (PASCAL) of the budget.
- **Strong effectiveness:** Achieves +27.23% (Cityscapes) and +14.30% (PASCAL) performance gains under the same budget.

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 ± 0.33	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

Table: Label correction is performed over five rounds with budgets of 10k/round (Cityscapes) and 1k/round (PASCAL).

Experiments: Qualitative results

- Constructed pseudo-labels



[1] REN, Tianhe, et al. Grounded sam: Assembling open-world models for diverse visual tasks. arXiv preprint arXiv:2401.14159, 2024.

[2] KIM, Hoyoung, et al. Active label correction for semantic segmentation with foundation models. In Forty-first International Conference on Machine Learning, 2024.

Experiments: Qualitative results

- Visualization of LCM-corrected masks

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



Experiments: Ablation study

Each component individually improves performance, while the best results are obtained when all are combined.

Notably, combining LCM and ABC yields greater gains than using either alone, highlighting the complementary and synergistic nature of these two components.

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

Table: Label correction is performed over five rounds with budgets of 10k/round (Cityscapes).

Conclusion

We introduce **A²LC**,

an Active and Automated Label Correction framework for semantic segmentation,
built upon cascading stages of manual and automatic correction.

The A²LC framework incorporates two key components:

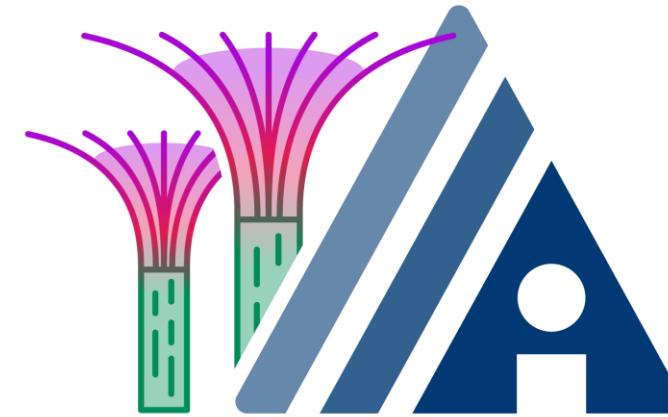
- (1) **LCM**, which performs automatic correction, and
- (2) **ABC**, which steers adaptively class-balanced correction.

Extensive experiments on Cityscapes and PASCAL VOC 2012 show that
our method achieves state-of-the-art performance with superior cost efficiency,
highlighting its potential for real-world deployment.

A²LC: Active and Automated Label Correction for Semantic Segmentation



Contact:
ngbrjyj@yonsei.ac.kr



AAAI-26 / IAAI-26 / EAAI-26
JANUARY 20-27, 2026 | SINGAPORE