Active Label Correction for Semantic Segmentation with Foundation Models

ICML 2024

Hoyoung Kim 1 Sehyun Hwang 2 Suha Kwak 1 2 Jungseul Ok 1 2

- **Problem/Objective**

  - label correction

- **Contribution/Key Idea**

  - correction query
  - look-ahead acquisition function

전유진

Active Label Correction for Semantic Segmentation with Foundation Models

ICML 2024

- **Pseudo code**

Algorithm 1 Proposed Framework

Require: Batch size B, and final round T.

Prepare initial dataset $\mathcal{D}_0$ requiring label correction
Obtain model θ_0 training with $\mathcal{D}_0$ via (1)
for $t = 1, 2, \ldots, T$ **do**
  Construct diversified pixel pool $\mathcal{X}_t^d$ via (4)
  Correct labels of selected B pixels $\mathcal{B}_t \subset \mathcal{X}_t^d$ via (9)
  Expand corrected labels to corresponding superpixels
  Obtain model θ_t training with corrected $\mathcal{D}_t$ via (11)
end for
return $\mathcal{D}_T$ and θ_T

전유진

Active Label Correction for Semantic Segmentation with Foundation Models

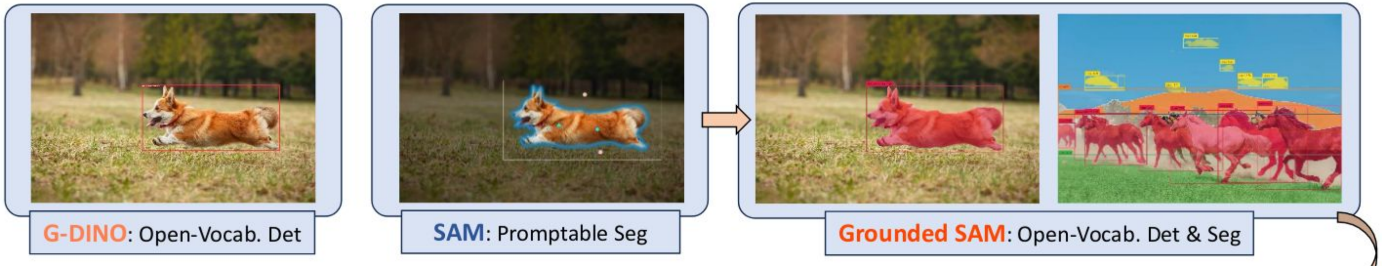
Algorithm 1 Proposed Framework

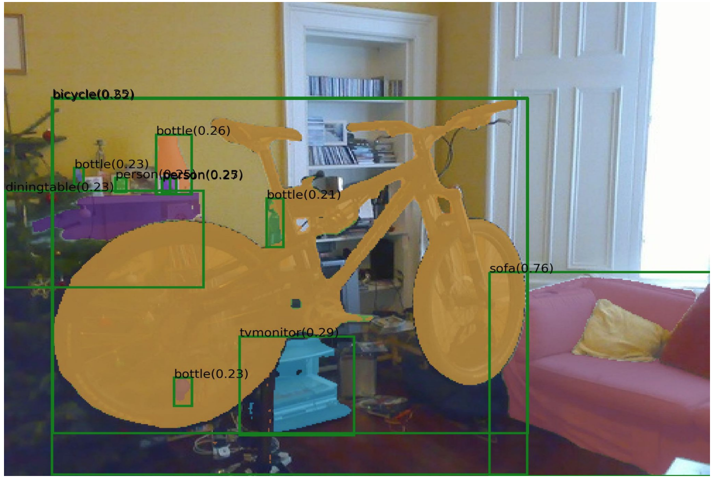
Require: Batch size B, and final round T.

1: Prepare initial dataset $\mathcal{D}_0$ requiring label correction
2: Obtain model θ_0 training with $\mathcal{D}_0$ via (1)
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Construct diversified pixel pool $\mathcal{X}_t^d$ via (4)
5: Correct labels of selected B pixels $\mathcal{B}_t \subset \mathcal{X}_t^d$ via (9)
6: Expand corrected labels to corresponding superpixels
7: Obtain model θ_t training with corrected $\mathcal{D}_t$ via (11)

- 1) Initial dataset preparation

Grounded SAM : detect and segment objects based on text prompts





$$\hat{\mathbb{E}}_{(x,y)\sim\mathcal{D}_0}\left[\mathrm{CE}(y, f_\theta(x))\right], \tag{1}$$

전유진

[Active Label Correction for Semantic Segmentation with Foundation Models]()

Algorithm 1 Proposed Framework

```
Require: Batch size B, and final round T.
  Prepare initial dataset D_0 requiring label correction
  Obtain model θ_0 training with D_0 via (1)
  for t = 1, 2, ..., T do
      Construct diversified pixel pool X_t^d via (4)
      Correct labels of selected B pixels B_t ⊂ X_t^d via (9)
      Expand corrected labels to corresponding superpixels
      Obtain model θ_t training with corrected D_t via (11)
  end for
  return D_T and θ_T
```

- **2) diversified pixel pool**

representative pixels per superpixels

$\mathcal{S}$: superpixel : organized based on the objects identified by G-SAM

$$\mathcal{X}^d := \{x_1, x_2, \ldots, x_{|\mathcal{S}|}\}\,, \qquad (3)$$

$$x_i := \underset{x \in s_i}{\arg\max} \frac{f_\theta(x) \cdot f_\theta(s_i')}{\|f_\theta(x)\| \|f_\theta(s_i')\|}\,, \qquad (4)$$

$$\mathrm{D}_\theta(s) := \underset{c \in \mathcal{C}}{\arg\max} \left|\{x \in s : y_\theta(x) = c\}\right|\,, \qquad (5)$$

$$s' := \{x \in s : y_\theta(x) = \mathrm{D}_\theta(s)\}\,. \qquad (6)$$

전유진

Algorithm 1 Proposed Framework

Require: Batch size B, and final round T.

Prepare initial dataset $\mathcal{D}_0$ requiring label correction
Obtain model θ_0 training with $\mathcal{D}_0$ via (1)
for $t = 1, 2, \ldots, T$ **do**
    Construct diversified pixel pool $\mathcal{X}_t^d$ via (4)
    Correct labels of selected B pixels $\mathcal{B}_t \subset \mathcal{X}_t^d$ via (9)
    Expand corrected labels to corresponding superpixels
    Obtain model θ_t training with corrected $\mathcal{D}_t$ via (11)
end for
return $\mathcal{D}_T$ and θ_T

- 3) look-ahead acquisition function

select the most informative pixels from the diversified pixel pool using acquisition function

$$x^* := \underset{x \in \mathcal{X}_t^d}{\arg\max}\, a(x; \theta_{t-1}) \,. \tag{7}$$

$$a_{\mathrm{CIL}}(x; \theta) := 1 - f_\theta(y; x) \,. \tag{8}$$

$$a_{\mathrm{SIM}}(x_r; s, \theta) := \sum_{x \in s} \frac{f_\theta(x_r) \cdot f_\theta(x)}{\|f_\theta(x_r)\| \|f_\theta(x)\|} a_{\mathrm{CIL}}(x; \theta) \,, \tag{9}$$

$$a_{\mathrm{LCIL}}(x_r; s, \theta) := \sum_{x \in s} \frac{1}{|s|} a_{\mathrm{CIL}}(x; \theta) \,. \tag{10}$$

[Active Label Correction for Semantic Segmentation with Foundation Models]

- 4) correction query

request the correct label when the given pseudo label is incorrect.

Algorithm 1 Proposed Framework

Require: Batch size B, and final round T.

Prepare initial dataset $\mathcal{D}_0$ requiring label correction
Obtain model θ_0 training with $\mathcal{D}_0$ via (1)
for $t = 1, 2, \ldots, T$ **do**
  Construct diversified pixel pool $\mathcal{X}_t^d$ via (4)
  Correct labels of selected B pixels $\mathcal{B}_t \subset \mathcal{X}_t^d$ via (9)
  Expand corrected labels to corresponding superpixels
  Obtain model θ_t training with corrected $\mathcal{D}_t$ via (11)
end for
return $\mathcal{D}_T$ and θ_T

Is this pixel a **boat**? Give the correct label only if the pseudo label is incorrect.

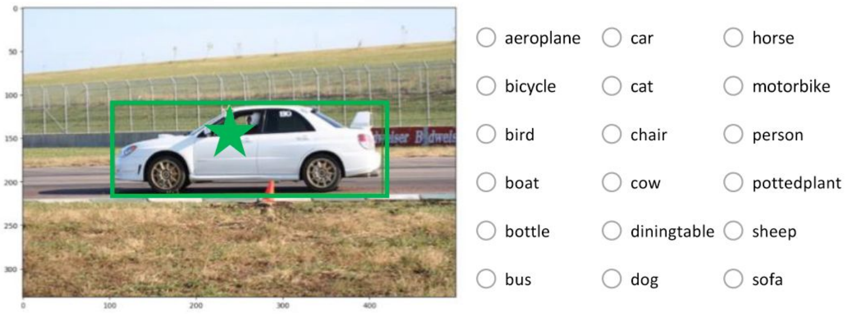
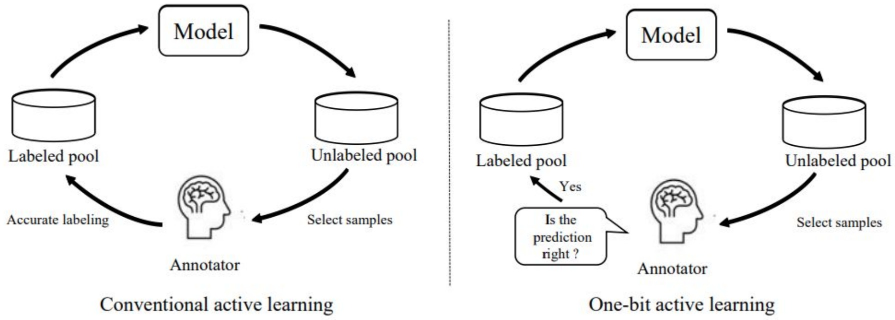


Figure 2: *An example of correction query*. Correction query presents an instruction requesting a label for a representative pixel (green star), an image displaying an object within a bounding box (green rectangle), and possible class options.

$$\hat{\mathbb{E}}_{(x,y)\sim\mathcal{D}_t}[\mathrm{CE}(y, f_\theta(x))] \ . \qquad (11)$$

참고) classification query   one-bit active query



전유진

- **Experiments**

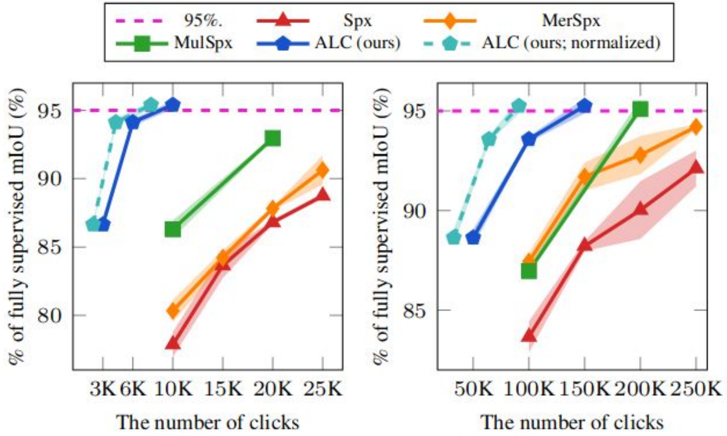
Table 1: *User study for different queries.* Our correction query C_{cor} proves to be more cost-effective compared to classification query C_{cls}.

Query	Total time (s)	Time per query (s)	Accuracy (%)
C_{cls}	$126.1_{\pm 19.8}$	$6.31_{\pm 0.99}$	$95.0_{\pm 3.3}$
C_{cor}	$\mathbf{95.1}_{\pm 9.0}$	$\mathbf{4.76}_{\pm 0.45}$	$95.0_{\pm 4.0}$

Table 2: *Quality of corrected datasets.* The labels of 5K pixels from the initial datasets are corrected using different acquisition functions in the ALC framework.

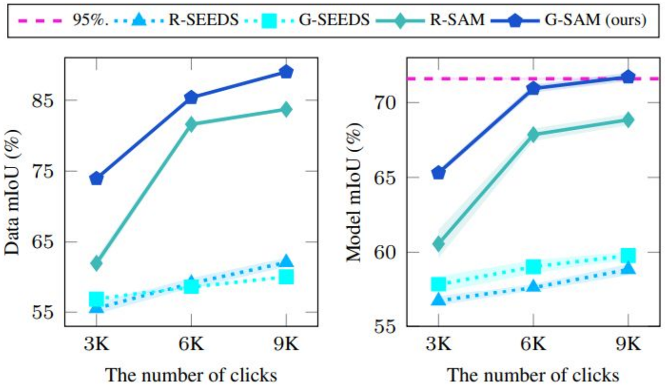
Acquisition function	Data mIoU (%)	Model mIoU (%)
LCIL	$56.59_{\pm 0.07}$	$56.82_{\pm 0.05}$
SoftMin	$59.28_{\pm 0.59}$	$58.66_{\pm 0.89}$
AIoU	$59.95_{\pm 0.57}$	$59.04_{\pm 0.27}$
SIM (ours)	$\mathbf{83.04}_{\pm 0.62}$	$\mathbf{68.72}_{\pm 0.10}$

- Experiments



(a) PASCAL (b) Cityscapes

Figure 3: *Effect of active label correction.* ALC shows comparable results on both datasets with much fewer clicks. ALC (normalized) reflects the reduced budget of correction queries with normalization by Theorem 3.1.



(a) Data mIoU (b) Model mIoU

Figure 6: *Advantages of foundation models.* Our ALC is called G-SAM, as it depends on Grounded-SAM. The effect of superpixels is larger than that of initial pseudo-labels.

전유진

- Experiments

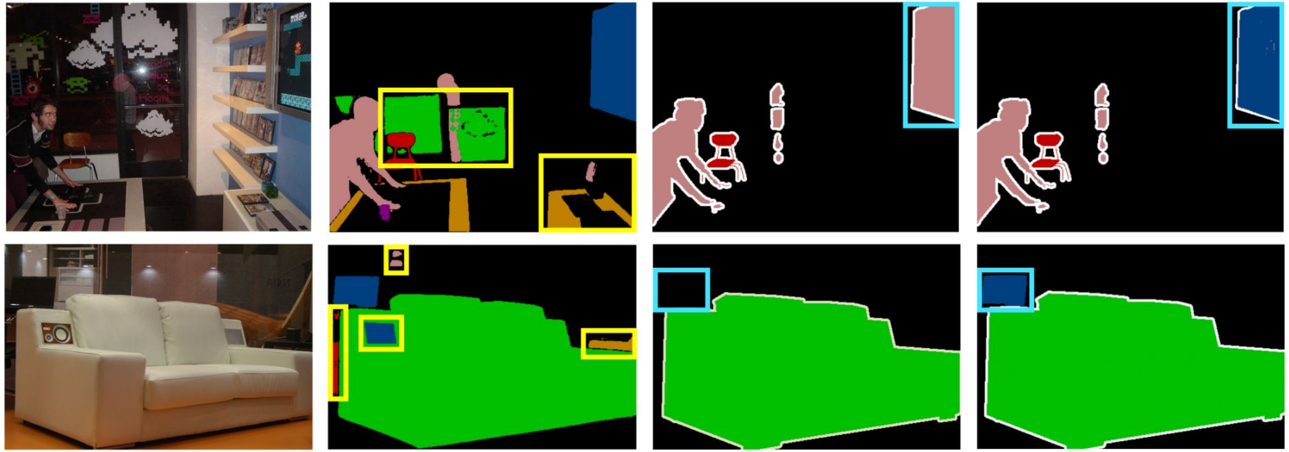
Table 3: *Synergy of proposed components.* We conduct an ablation study, when correcting the initial dataset using 5K budgets in PASCAL.

Acquisition: Diversity	Acquisition: Look-ahead	Expansion	Data mIoU	Model mIoU
✗	✗	✗	$55.03_{\pm 0.25}$	$56.30_{\pm 0.56}$
✗	✓	✓	$55.38_{\pm 0.08}$	$56.01_{\pm 0.58}$
✓	✗	✓	$56.59_{\pm 0.07}$	$56.82_{\pm 0.05}$
✓	✓	✗	$55.61_{\pm 0.00}$	$56.69_{\pm 0.35}$
✓	✓	✓	$\mathbf{83.04}_{\pm 0.62}$	$\mathbf{68.72}_{\pm 0.10}$

Table 4: *Fair comparison between Spx and ALC.* For a fair comparison, we integrate two advantages of foundation models into Spx. We refine the initial dataset using 3K budgets in PASCAL.

Methods	Initial stage	Superpixels	Model mIoU (%)
Spx	Cold-start	SEEDS	$52.34_{\pm 0.85}$
Spx	Warm-start	SEEDS	$57.77_{\pm 0.70}$
Spx	Warm-start	SAM	$57.79_{\pm 0.66}$
ALC	Warm-start	SAM	$\mathbf{65.30}_{\pm 0.21}$

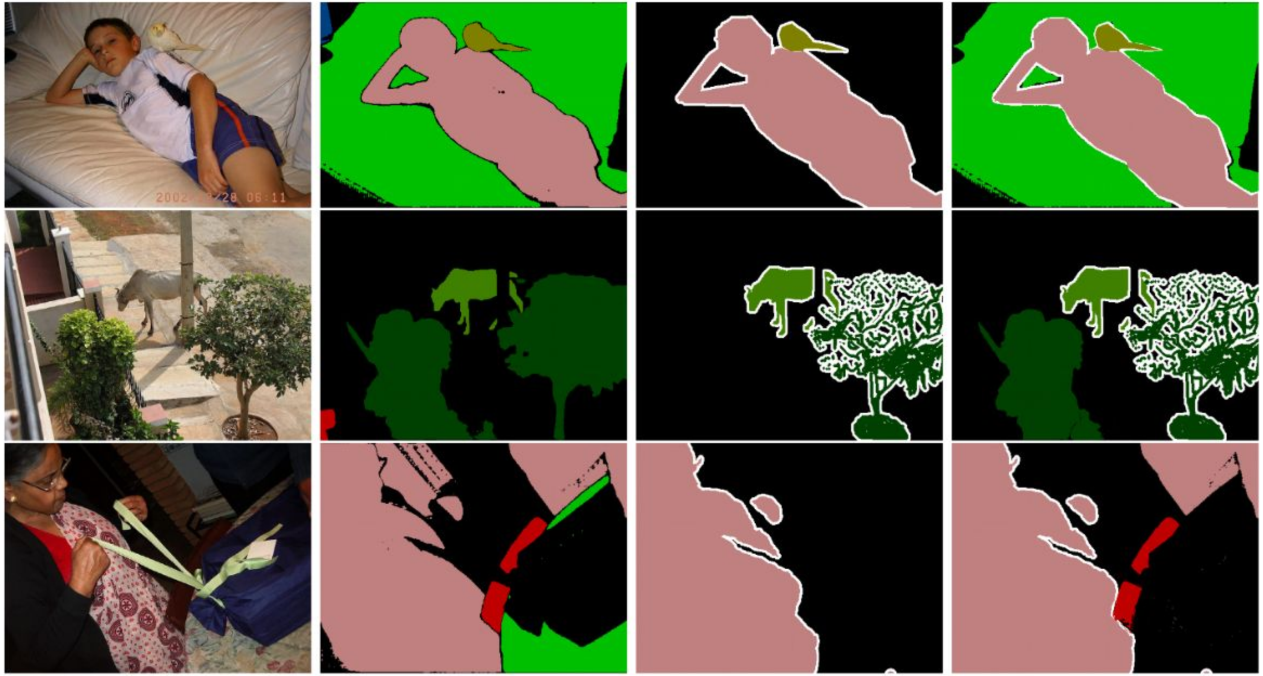
- Result



(a) Unlabeled image (b) Grounded-SAM (c) PASCAL (d) PASCAL+ (ours)

Figure 1: *Examples of noisy and corrected labels in PASCAL.* (a, b) Initial pseudo labels are generated by applying Grounded-SAM (G-SAM) to unlabeled images. As depicted by the yellow boxes, noisy pseudo labels result in a decline in performance, as shown in Table 7. (c) PASCAL also contains noisy labels in cyan boxes. (d) By employing the superpixels from G-SAM, we construct a corrected version of PASCAL, called PASCAL+. For instance, in the first row, we correct the object labeled as person to tvmonitor, and in the second row, the object labeled as background to tvmonitor. Here, the colors black, blue, red, green, and pink represent the background, tvmonitor, chair, sofa, and person classes, respectively.

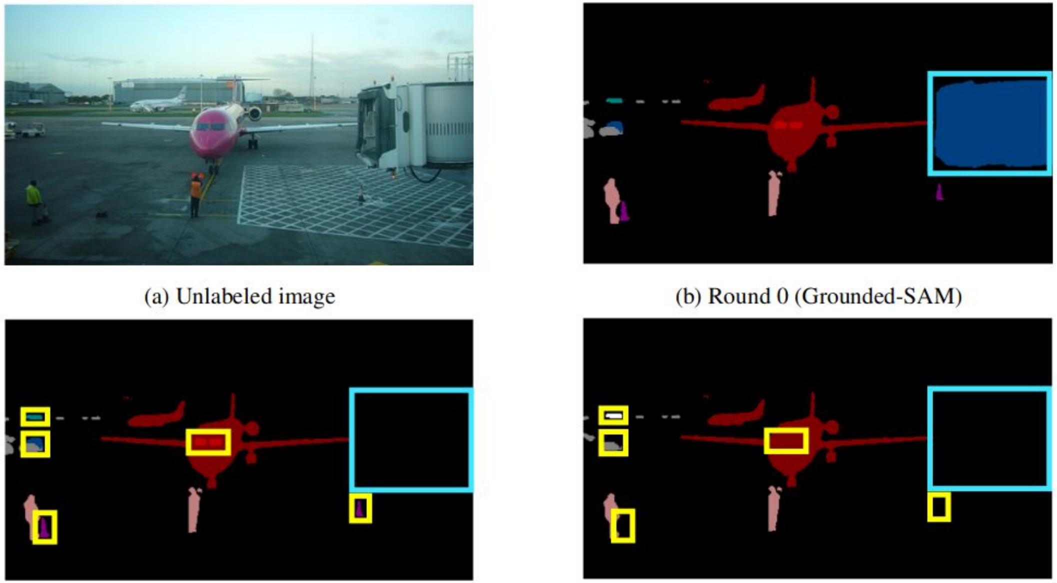
- **Result**



(a) Unlabeled image (b) Grounded-SAM (c) PASCAL (d) PASCAL+ (ours)

Figure 9: *Additional examples of noisy and corrected labels in PASCAL.* We correct PASCAL into PASCAL+ utilizing the superpixels of Grounded-SAM.

- **Result**



(a) Unlabeled image

(b) Round 0 (Grounded-SAM)

(c) Round 1

(d) Round 2

Figure 13: *Segmentation changes through active label correction.* (b) The initial pseudo labels obtained from Grounded-SAM contain numerous noisy labels, exemplified by instances like tvmonior inside the cyan box. (c) In the first round, the object labeled as tvmonitor is corrected to background. Nonetheless, many noisy labels exist within the yellow boxes. (c) In the second round, we rectify all remaining noisy labels. With the help of the proposed look-ahead acquisition function, we prioritize correcting large objects before addressing small ones. Here, the colors black, blue, red, dark red, purple, and pink represent the background, tvmonitor, chair, airplane, bottle and person classes, respectively.

전유진