논문 제목: A^2LC: Active and Automated Label Correction for Semantic Segmentation

전유진

ArXiv preprint posted: June 13, 2025

This presentation summarizes the baselines of our paper.

주제 및 Contribution	• Semantic Segmentation
저자 별 연구 내용	• Label Correction Module (LCM) • Adaptively Balanced Acquisition Function (ABC)
기타	• Jeon, Youjin, et al. "A $^ 2$ LC: Active and Automated Label Correction for Semantic Segmentation." *arXiv preprint arXiv:2506.11599* (2025).

Revisiting Superpixels for Active Learning in Semantic Segmentation with Realistic Annotation Costs

Lile Cai[1], Xun Xu[1], Jun Hao Liew[2], Chuan Sheng Foo[1]
[1]Institute for Infocomm Research, Singapore
[2]National University of Singapore
{caill, foo_chuan_sheng}@i2r.a-star.edu.sg, alex.xun.xu@gmail.com, liewjunhao@u.nus.edu

- **Problem / objective**
  - Semantic Segmentation에서, 픽셀 단위 라벨링의 높은 난이도로 인해, 데이터셋 확보 어려움

- **Contribution / Key idea**
  - Region-based Active Learning (AL)
    - 라벨링 단위: (Irregulary-shaped) Superpixels
    - 라벨링 비용 책정 단위: Click
    - Class-balanced acquisition function

Methods

- **Overall Framework**

1. 이미지를 슈퍼픽셀들로 분할.
2. 정보량 많은 데이터를 class-balanced 샘플링.
3. Oracle에게 레이블 요청. (실험에서는 gt 라벨 사용.)
   a. 기존의 polygon-based labeling 대신에, dominant labeling 수행
   b. Dominant labeling: 각 슈퍼픽셀마다 하나의 레이블 할당.
4. 현재까지 라벨링된 데이터들 사용하여 모델 재학습.
5. 그리고, 위 과정은 라벨링 예산 소진될때까지 반복.

Cai, Lile, et al. "Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Methods

- **Superpixel Generation**

SEEDS 알고리즘 사용 [1]

Fig. 14 Example SEEDS segmentations with 200 superpixels. The ground-truth segments are color coded and blended on the images. The superpixel boundaries are shown in white.

[1] Van den Bergh, Michael, et al. "Seeds: Superpixels extracted via energy-driven sampling." *International Journal of Computer Vision* 111.3 (2015): 298-314.
[2] Cai, Lile, et al. "Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Methods

- **Class-Balanced Sampling**

  - **AL에서 그다음에 쿼리할 샘플**

$$s^* = \arg\max_{s \in \mathcal{U}_t} a(s, M_t). \qquad (1)$$

  - 픽셀 x 의 불확실성 (BvSB)

$$u(x, M_t) = \frac{p(y = c^{sb} | x, M_t)}{p(y = c^b | x, M_t)}, \qquad (2)$$

  - 슈퍼픽셀 s 의 불확실성 **(BvSB)**

$$u(s, M_t) = \frac{\sum_{x \in s} u(x, M_t)}{|\{x : x \in s\}|}. \qquad (3)$$

  - 슈퍼픽셀 s 의 레이블

$$\begin{aligned} Do(s) &= \arg\max_{c \in \mathcal{C}} |\{x : l(x) == c \text{ and } x \in s\}|, \\ l(x) &= \arg\max_{c \in \mathcal{C}} p(y = c | x, M_t), \end{aligned} \qquad (4)$$

Cai, Lile, et al. "Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Methods

● Class-Balanced Sampling

❑ 슈퍼픽셀 s 의 클래스 분포

$$p(cls) = \frac{|\{s : Do(s) == cls\}|}{\sum_{c \in \mathcal{C}} |\{s : Do(s) == c\}|}. \qquad (5)$$

❑ **최종 획득함수 (ClassBal)**

$$a(s, M_t) = u(s, M_t) e^{-p(Do(s))}. \qquad (6)$$

❑ **획득함수를 통한 샘플링 과정**

Algorithm 1: Batch-Mode Active Selection

Input : unlabeled set of regions $\mathcal{U}_t$, labeled set of regions $\mathcal{L}_{t-1}$ selected in previous batches, model M_t trained on $\mathcal{L}_{t-1}$, annotation budget of K clicks for batch t

Output: Output selected set of regions $\mathcal{B}_t$

$\mathcal{B}_t = \emptyset$;

$total\_cost = 0$;

while $total\_cost < K$ **do**

  $s^* = \arg\max_{s \in \mathcal{U}_t} a(s, M_t)$;

  $\mathcal{B}_t = \mathcal{B}_t \cup s^*$;

  $\mathcal{U}_t = \mathcal{U}_t \setminus s^*$;

  $total\_cost = total\_cost + cost(s^*)$;

end

Cai, Lile, et al. "Revisiting superpixels for active learning in se ... of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

김유진

Methods

- **Annotation Cost Measurement**

  - **라벨링 비용**
    a. 클릭 개수 기반 [1] (cf., 기존: 라벨링된 픽셀 개수 기반)

  - 클릭 종류
    - Polygon clicks (c_p): Polygon 꼭짓점 개수
    - Class clicks (c_c): 각 polygon 내에 connected component 개수
    - Intersection clicks (c_i): 지역 경계와 객체 경계 교차점 개수 (in "rectangle+polygon" 접근법)

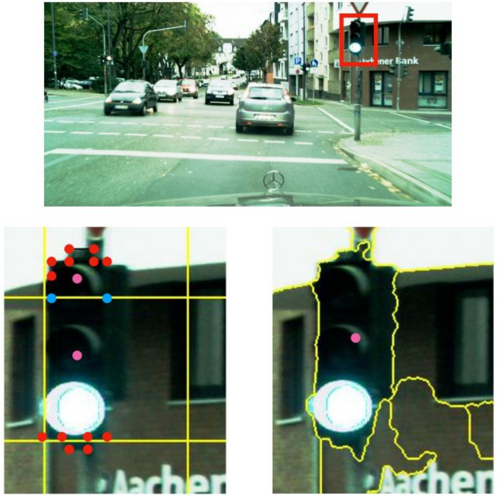

Figure 1: Annotating a traffic light by "rectangle+polygon" based approach (bottom left) *vs*. the superpixel-based approach (bottom right). The former requires quite a few polygon clicks (red dots), intersection clicks (blue dots) and class clicks (pink points), while the latter only requires a class click. If the annotation cost is measured in pixels, the two schemes perform closely, yet when measured in clicks, the latter is much more efficient.

[1] Colling, Pascal, et a… …segmentation using priority maps." *arXiv preprint arXiv:2010.01884* (2020).
[2] Cai, Lile, et al. "Rev… …alistic annotation costs." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Methods

- **Annotation Cost Measurement**

  - **라벨링 방법**
    a. 슈퍼픽셀 기반 Dominant labeling (cf., 기존: "rectangle+polygon" 기반 Precise labeling)

  - 라벨링 방법 종류
    - Precise labeling (Pr): c_p + c_c + c_i
    - Dominant labeling (Do): c_c

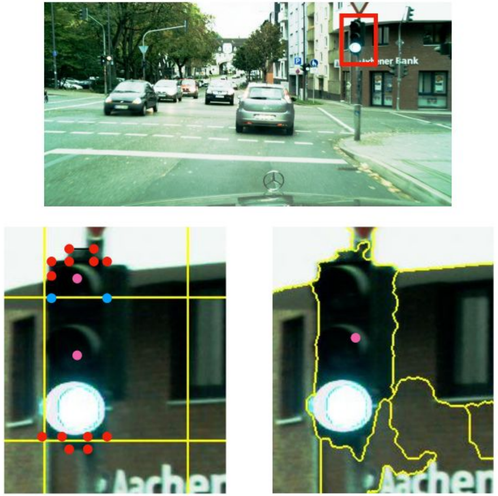

Figure 1: Annotating a traffic light by "rectangle+polygon" based approach (bottom left) *vs*. the superpixel-based approach (bottom right). The former requires quite a few polygon clicks (red dots), intersection clicks (blue dots) and class clicks (pink points), while the latter only requires a class click. If the annotation cost is measured in pixels, the two schemes perform closely, yet when measured in clicks, the latter is much more efficient.

[1] Colling, Pascal, et a … segmentation using priority maps." *arXiv preprint arXiv:2010.01884* (2020).
[2] Cai, Lile, et al. "Revisiting superpixels for active learning in semantic segmentation with realistic annotation costs." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

Adaptive Superpixel for Active Learning in Semantic Segmentation

Hoyoung Kim[1] Minhyeon Oh[2] Sehyun Hwang[2] Suha Kwak[1,2] Jungseul Ok[1,2*]

Graduate School of AI, POSTECH[1], Dept. of CSE, POSTECH[2]
{cskhy16, minhyeonoh, sehyun03, suha.kwak, jungseul}@postech.ac.kr

- **Problem / objective**
  - Semantic Segmentation에서, 픽셀 단위 라벨링의 높은 난이도로 인해, 데이터셋 확보 어려움
- **Contribution / Key idea**
  - Superpixel-based AL framework
    - Adaptive Merging mechanism
    - Sieving mechanism

Introduction

- **Motivation**

1. Region-based AL 방법론에서, "쿼리할 region"의 역할 중요
    a. Region이 너무 작으면, 라벨링 효율 감소
    b. Region이 너무 크면, 노이즈 증가

2. 슈퍼픽셀 생성 알고리즘들을 통해 생성된 "슈퍼픽셀"에는 제약이 많이 걸려있음
    a. ∵ 비슷한 크기 및 모양들로 색상 비슷한 인접 픽셀들 클러스터링하는 원리

3. **따라서, "슈퍼픽셀 개선" 방법론 포함한 AL 프레임워크 제안**
    a. Adaptive Merging → 라벨링 효율 증가
    b. Sieving → 노이즈 감소

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

- **Overview**

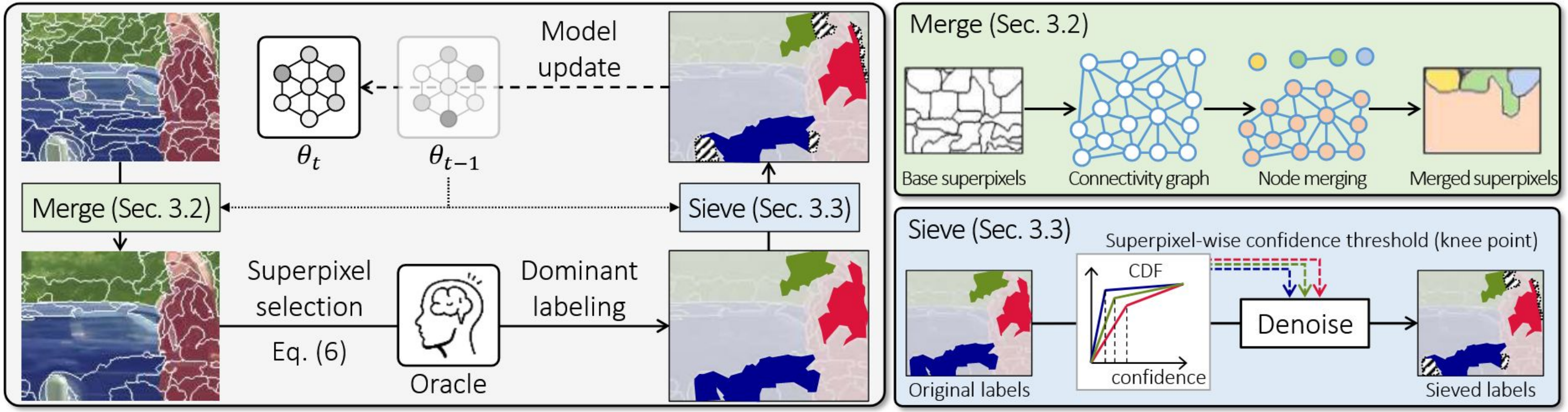

Figure 2: *An overview of the proposed framework.* In each round t, we merge superpixels with a graph using the latest model, and obtain dominant labels for selected superpixels. The dominant labels are selectively propagated to pixels with confidence above the detected knee point, resulting in the creation of a sieved dataset. Finally, we train a model with the sieved one.

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

- **Overview**

Algorithm 1 Proposed Framework

Require: Image set $\mathcal{I}$, batch size B, and final round T.

1: Produce base superpixels $\mathcal{S}_0 := \bigcup_{i \in \mathcal{I}} \mathcal{S}_0(i)$
2: Obtain model θ_0 training with $\mathcal{D}_0$
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Adaptively merge the base superpixels and obtain $\mathcal{S}_t \leftarrow \bigcup_{i \in \mathcal{I}} \mathrm{AM}(\mathcal{S}_0(i), \theta_{t-1})$
5: Select and query B superpixels $\mathcal{B}_t \subset \mathcal{S}_t$ with (7)
6: Sieve $s \in \bigcup_{t'=0}^{t} \mathcal{B}_{t'}$ and obtain $\mathcal{D}_t$ in (9)
7: Obtain model θ_t training with the sieved $\mathcal{D}_t$
8: **return** θ_T

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

- **Warm-up round**

❑ **SEEDS 알고리즘을 통해, 슈퍼픽셀 집합 (Base Superpixels) 생성**

$$\mathcal{S}_0 := \bigcup_{i \in \mathcal{I}} \mathcal{S}_0(i)$$

❑ 쿼리할 B개의 슈퍼픽셀 랜덤 샘플링

$$\mathcal{B}_0$$

❑ 쿼리해서 레이블 획득 (Dominant labeling)

$$\mathcal{D}_0 := \{(x, y) : \exists s \in \mathcal{B}_0, x \in s, y(c) = \mathbb{1}[c = \mathrm{D}(s)] \ \ \forall c \in \mathcal{C}\}$$

❑ 획득한 라벨로 모델 학습

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

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

- **Adaptive merging**

  - Algorithm

Algorithm 2 Adaptive Merging (AM)

Require: Base superpixels S, model θ, and threshold ϵ.

Set $S' \leftarrow \emptyset$ and $\mathcal{G}(S) \leftarrow (S, \mathcal{E}(S))$

Mark s as unexplored for each $s \in S$

for $s \in S$ in descending order of $u_\theta(s)$ **do**

    if s is unexplored **then**

        $S' \leftarrow S' \cup \{\textsc{Merge}(s, f_\theta(s); \mathcal{G}, \theta)\}$

return S'

procedure $\textsc{Merge}(s, f; \mathcal{G}, \theta)$

    Mark s as explored and set $s' \leftarrow s$

    for each neighbor n of s in $\mathcal{G}$ **do**

        if n is unexplored and $d_{\mathrm{JS}}(f \,\|\, f_\theta(n)) < \epsilon$ **then**

            $s' \leftarrow s' \cup \textsc{Merge}(n, f; \mathcal{G}, \theta)$

    return s'

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

● Adaptive merging

❑ 각 슈퍼픽셀들을 노드로 그래프 연결

$$\mathcal{G}(S) = (S, \mathcal{E}(S))$$

❑ **인접 노드들을 합병한 결과, 기존 슈퍼픽셀 집합 기반으로 합병된 새로운 슈퍼픽셀 집합 (Merged Superpixels) 획득**

a. 탐색 방법: Breadth First Search (BFS) 알고리즘 사용

b. 거리 메트릭: Jensen-Shannon (JS) divergence 사용

$$d_{\mathrm{JS}}\big(f_\theta(s) \parallel f_\theta(n)\big) < \epsilon \,, \qquad (2)$$

$$d_{\mathrm{JS}}(p \parallel q) := \sqrt{\frac{d_{\mathrm{KL}}(p \parallel \frac{p+q}{2}) + d_{\mathrm{KL}}(q \parallel \frac{p+q}{2})}{2}} \,, \qquad (3)$$

c. 합병된 슈퍼픽셀 집합 (Merged Superpixels)

$$\mathcal{S}_t := \bigcup_{i \in \mathcal{I}} S_t(i)$$

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

YONSEI University CI Computational Intelligence Laboratory

- **Adaptive merging**

❑ 픽셀 x의 불확실성 (BvSB)

$$u_\theta(x) := \frac{\max_{c\in\mathcal{C}\setminus\{y_\theta(x)\}} f_\theta(c;x)}{\max_{c\in\mathcal{C}} f_\theta(c;x)}\,, \qquad (4)$$

❑ 슈퍼픽셀 s의 불확실성 (BvSB)

$$u_\theta(s) := \frac{\sum_{x\in s} u_\theta(x)}{|\{x : x\in s\}|}\,, \qquad (5)$$

❑ 픽셀 x의 클래스 분포

$$p(c;\theta) := \frac{|\{x : \exists s\in\mathcal{S}_t, \mathrm{D}_\theta(s)=c, x\in s\}|}{|\{x : \exists s\in\mathcal{S}_t, x\in s\}|}\,, \qquad (6)$$

❑ 최종 획득함수

$$a(s;\theta) := u_\theta(s)\exp\left(-p(\mathrm{D}_\theta(s);\theta)\right). \qquad (7)$$

❑ **합병된 슈퍼픽셀 집합 (Merged Superpixels) 에서 쿼리할 B개의 슈퍼픽셀 샘플링**

$$\mathcal{B}_t$$

전유진

Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

Proposed framework

- **Sieving**

  - Sieve 원리: Kneedle 알고리즘 [1] 을 통해 각 슈퍼픽셀 내 픽셀들의 모델 예측 확률 누적 분포에서 knee point를 탐색 및 해당 지점 이상의 픽셀들을 제거

$$h(s;\theta) := \{x \in s : f_\theta(\mathrm{D}(s);x) \geq \phi(s;\theta)\}\,, \qquad (8)$$

  - 현재 라운드까지 쿼리된 모든 슈퍼픽셀 집합에 대해 sieve 과정을 적용하여 새로운 sieved dataset을 구축

$$\mathcal{D}_t := \left\{ (x,y) : \begin{matrix} \exists s \in \cup_{t'=0}^{t} \mathcal{B}_{t'},\ x \in h(s;\theta_{t-1}), \\ y(c) = \mathbb{1}[c = \mathrm{D}(s)]\ \forall c \in \mathcal{C} \end{matrix} \right\}. \quad (9)$$

  - 모델 학습

$$\hat{\mathbb{E}}_{(x,y)\sim\mathcal{D}_t}[\mathrm{CE}(y, f_\theta(x))]\,. \qquad (10)$$

[1] Satopaa, Ville, et al. "Finding a" kneedle" in a haystack: Detecting knee points in system behavior." *2011 31st international conference on distributed computing systems workshops*. IEEE, 2011.
[2] Kim, Hoyoung, et al. "Adaptive superpixel for active learning in semantic segmentation." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

전유진

YONSEI University CI Computational Intelligence Laboratory

Active Learning for Semantic Segmentation with Multi-class Label Query

Sehyun Hwang Sohyun Lee Hoyoung Kim Minhyeon Oh Jungseul Ok Suha Kwak

Pohang University of Science and Technology (POSTECH), South Korea
{sehyun03, lshig96, cskhy16, minhyeonoh, jungseul.ok, suha.kwak}@postech.ac.kr

- **Problem / objective**
  - Semantic Segmentation에서, 픽셀 단위 라벨링의 높은 난이도로 인해, 데이터셋 확보 어려움
- **Contribution / Key idea**
  - Region-based AL framework
    - Region-wise multi-class labeling strategy
    - Two-stage training algorithm using partial labels

Introduction

- **Motivation**

1. AL에서 "라벨링 쿼리 디자인" 은 중요
    a. ∵ 각 쿼리가 라벨링 예산과 직결되기 때문

2. **쿼리 디자인으로, "Region-based Multi-class label query" 전략 새롭게 제안**
    a. 쿼리 단위: "Region-based"
        i. 이미지 단위로 쿼리 시, 샘플의 다양성 부족
        ii. 픽셀 단위로 쿼리 시, 예산의 비효율적 사용
        iii. ∴ Non-overlapped local image region 단위로 쿼리해서, 이미지단과 픽셀단의 절충안 사용
    b. 라벨링 방법: "Multi-class labeling"
        i. "Multi-class labeling" 이 "Dominant-class labeling" 보다 아래 두가지 이유로 더 나음
            - ❏ 클릭당 라벨링 소요 시간↓
            - ❏ 더 정확한 레이블링↑
        ii. Figure 1 참조

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Introduction

- **Motivation**

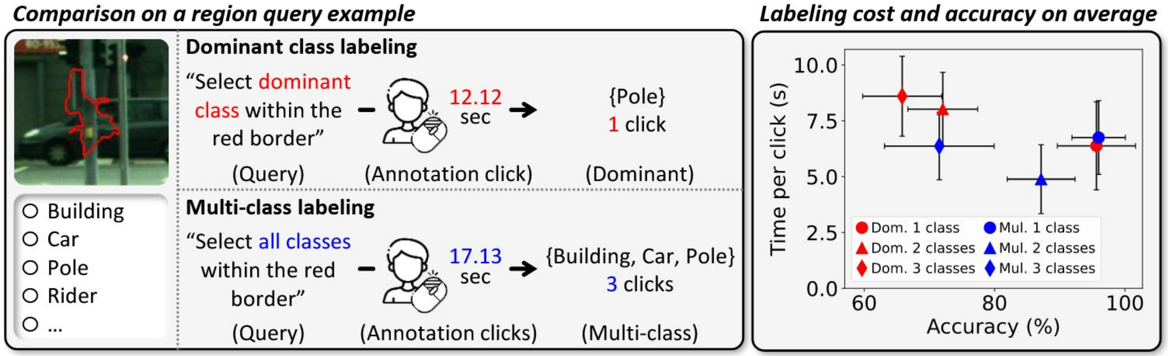

Figure 1: Dominant class labeling [9] versus our multi-class labeling. (*left*) Given a local region as query, an oracle is asked to select the most dominant class by a single click in dominant class labeling, and all existing classes by potentially more than one click in multi-class labeling. As shown here, multi-class labeling often takes less *annotation time per click* because, to determine the dominant one, the oracle has to infer every class in the region after all and sometimes should very carefully investigate the region when the classes occupy areas of similar sizes. (*right*) We conducted a user study to compare the two strategies in terms of actual labeling cost and accuracy versus the number of classes in region queries; the results are summarized in the right plot with one standard deviation. Multi-class labeling took less time per click on average due to the above reason. Furthermore, it resulted in more accurate labels by annotating non-dominant classes ignored in dominant class labeling additionally. Details of this user study are given in Appendix A.

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Introduction

- **Motivation**

1. **앞서 제안한 Multi-class label query 전략에 적합한 "두 단계의 학습 알고리즘" 제안**
   a. 문제점: "Multi-class label query" 전략의 "Class ambiguity" 이슈
      i. 픽셀마다 partial labels 만 주어져 있어, 어떤 클래스가 실제 정답인지 불확실한 클래스 모호성 문제로 인해 학습 불안정
   b. 해결책: "Two-stage training algorithm" 제안
      i. Stage 1: Learning with region-wise multi-class labels
      ii. Stage 2: Learning with pixel-wise pseudo labels

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Proposed Method

- **Overview**

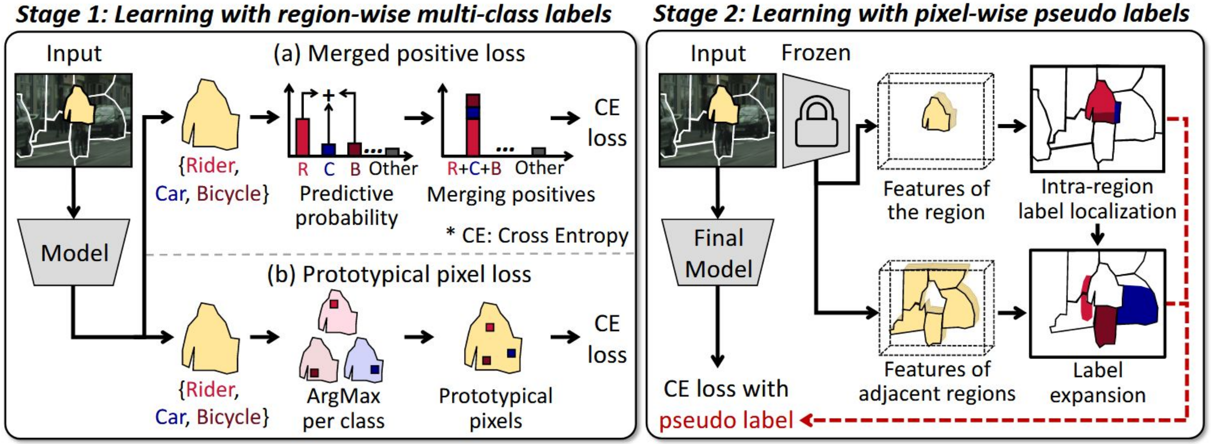

Figure 2: Our two-stage training algorithm using partial labels. (*left*) In the first stage, a model is trained using region-wise multi-class labels through two losses: the merged positive loss that encourages the model to predict any of the annotated classes for each pixel of the region, and the prototypical pixel loss that ensures at least one pixel in the region corresponds to each annotated class. (*right*) The second stage disambiguates the region-wise multi-class labels by generating pixel-wise pseudo labels, which are then used for training the final model. To this end, it first assigns pseudo class labels to individual pixels within the region (*i.e.*, intra-region label localization), and then propagates the pseudo labels to adjacent regions (*i.e.*, label expansion).

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Proposed Method

- **Acquisition of region-wise multi-class labels**

❑ 이미지

$$\mathcal{I}$$

❑ Non-overlapping regions

$$\mathcal{S} := \bigcup_{I \in \mathcal{I}} S(I)$$

❑ B개의 regions 쿼리

$$\mathcal{B}_t \subset \mathcal{S}$$

❑ **Annotator로부터 multi-class labels 획득**

$$\mathcal{D}_t := \{(s, Y) : s \in \mathcal{B}_t\} \quad Y \subset \{1, 2, \cdots, C\}$$

❑ 현재 라운드 t 까지 얻은 레이블들로 모델 학습

$$\theta_t \qquad \mathcal{D} := \bigcup_t \mathcal{D}_t$$

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

- **Acquisition of region-wise multi-class labels**

❑ 픽셀 x가 클래스 c에 속할 확률

$$P_{\theta_t}(y = c|x) = \mathrm{softmax}\left(\frac{f_t(x)^\top \mathbf{w}_{t,c}}{\tau\|f_t(x)\|\,\|\mathbf{w}_{t,c}\|}\right), \quad (1)$$

❑ 픽셀 x의 불확실성 (BvSB)

$$u_{\theta_t}(x) := \frac{P_{\theta_t}(y = c_{\mathrm{sb}}|x)}{P_{\theta_t}(y = c_{\mathrm{b}}|x)}, \quad (2)$$

❑ 픽셀 x의 클래스 분포

$$P_{\theta_t}(y = c) = \frac{1}{|X|}\sum_{x\in X} P_{\theta_t}(y = c|x), \quad (3)$$

❑ **최종 획득함수**

$$a(s;\theta_t) := \frac{1}{|s|}\sum_{x\in s} \frac{u_{\theta_t}(x)}{\left(1 + \nu\, P_{\theta_t}(c_b)\right)^2}, \quad (4)$$

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Proposed Method

- **Stage 1: Learning with region-wise multi-class labels**

❑ Single class로 레이블링된 regions

$$\mathcal{D}_{\mathrm{s}} := \left\{(s, \{c\}) : \exists (s, Y) \in \mathcal{D},\ |Y| = 1,\ c \in Y\right\}. \quad (5)$$

❑ Cross-Entropy (CE) loss

$$\mathcal{L}_{\mathrm{CE}} = \hat{\mathbb{E}}_{(s,\{c\})\sim\mathcal{D}_{\mathrm{s}}}\left[\frac{1}{|s|}\sum_{x\in s} -\log P_\theta(y = c|x)\right]. \quad (6)$$

❑ Multple classes로 레이블링된 regions

$$\mathcal{D}_{\mathrm{m}} := \mathcal{D} - \mathcal{D}_{\mathrm{s}}$$

❑ **Merged positive loss: 각 region에 속하는 픽셀들이 레이블 후보군 안에서 예측되도록 유도**

$$\mathcal{L}_{\mathrm{MP}} := \hat{\mathbb{E}}_{(s,Y)\sim\mathcal{D}_{\mathrm{m}}}\left[\frac{1}{|s|}\sum_{x\in s} -\log \sum_{c\in Y} P_\theta(y = c|x)\right]. \quad (7)$$

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

- **Stage 1: Learning with region-wise multi-class labels**

❑ Prototypical pixel: 각 region 내에서 레이블 후보 클래스별로 가장 대표적인 픽셀

$$x^*_{s,c} := \max_{x \in s} P_\theta(y = c|x) \;, \qquad (8)$$

❑ **Prototypical pixel loss: Prototypical pixel에 대해서는 정답 레이블일 것이라고 가정하고 학습**

$$\mathcal{L}_{\mathrm{PP}} := \hat{\mathbb{E}}_{(s,Y)\sim\mathcal{D}_{\mathrm{m}}}\left[\frac{1}{|Y|}\sum_{c\in Y} -\log P_\theta(y = c|x^*_{s,c})\right] . \qquad (9)$$

❑ 최종 loss

$$\mathcal{L} = \lambda_{\mathrm{CE}}\, \mathcal{L}_{\mathrm{CE}} + \lambda_{\mathrm{MP}}\, \mathcal{L}_{\mathrm{MP}} + \mathcal{L}_{\mathrm{PP}} \;, \qquad (10)$$

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Proposed Method

- **Stage 2: Learning with pixel-wise pseudo labels**

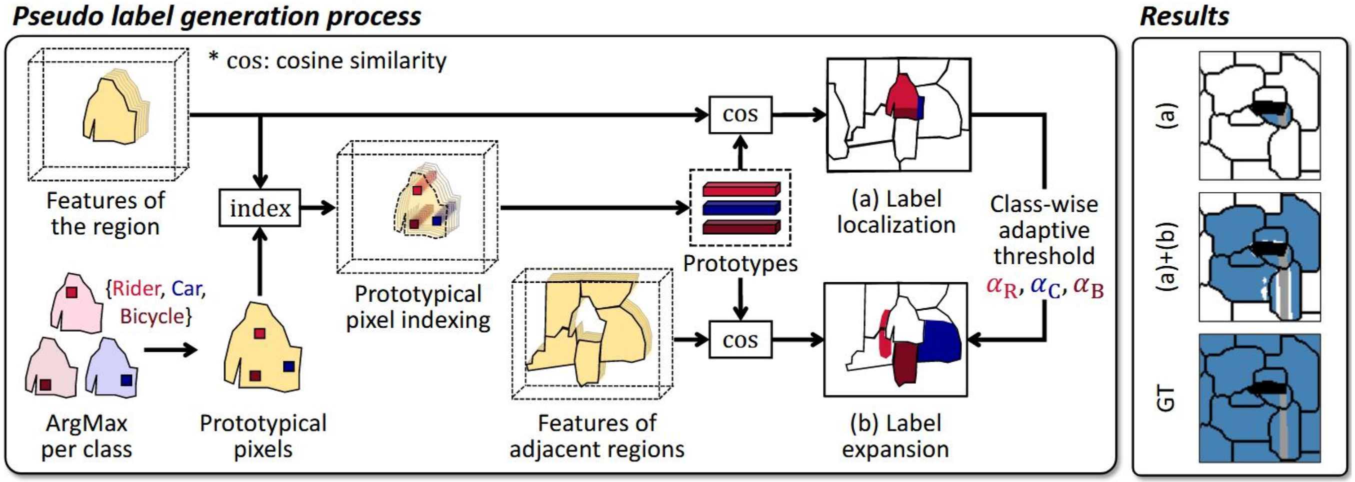

Figure 3: The pseudo label generation process (*left*) and its qualitative results (*right*). In each of the labeled regions, the feature vector located at the prototypical pixel of an annotated class is considered the prototype of the class, and the set of such prototypes is used as a region-adaptive classifier for pixel-wise pseudo labeling within the region (label localization). The pseudo labels of the region are propagated to adjacent unlabeled regions similarly (label expansion), but for conservative propagation, only relevant pixels that are close to at least one prototype will be assigned pseudo labels.

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

Proposed Method

YONSEI University Computational Intelligence Laboratory

- **Stage 2: Learning with pixel-wise pseudo labels**

1. **Intra-region label localization**
   a. Region 내에 class prototype과 유사한 픽셀들에 해당 클래스로 레이블링
   b. Region 내 픽셀 x에 할당될 레이블

$$\hat{y}(x) := \arg\max_{c \in Y} \cos\big(f_\theta(x), f_\theta(x^*_{s,c})\big)\ , \tag{11}$$

2. **Label expansion**
   a. 해당 region과 인접한 region들에도 유사도 기반으로 레이블 할당
   b. Class prototype adaptive threshold

$$\alpha_c(s) = \text{med}\Big(\big\{\cos\big(f_\theta(x), f_\theta(x^*_{s,c})\big) : x \in s,\ \hat{y}(x) = c\big\}\Big)\ , \tag{12}$$

   c. 위 threshold 넘으면 인접 픽셀로 간주

$$\hat{Y}(x) := \big\{c : \cos\big(f_\theta(x), f_\theta(x^*_{s,c})\big) > \alpha_c(s),\ c \in Y\big\} \qquad |\hat{Y}(x)| \geq 1$$

   d. Region 밖 인접 픽셀 x에 할당될 레이블

$$\hat{y}(x) := \arg\max_{c \in \hat{Y}(x)}\ \cos\big(f_\theta(x), f_\theta(x^*_{s,c})\big) \quad \text{only if } |\hat{Y}(x)| \geq 1\ , \tag{13}$$

전유진

Hwang, Sehyun, et al. "Active learning for semantic segmentation with multi-class label query." *Advances in Neural Information Processing Systems* 36 (2023): 27020-27039.

YONSEI University CI Computational Intelligence Laboratory

Active Label Correction for Semantic Segmentation with Foundation Models

Hoyoung Kim [1] **Sehyun Hwang** [2] **Suha Kwak** [1 2] **Jungseul Ok** [1 2]

- **Problem / objective**
  - Semantic Segmentation에서, 픽셀 단위 라벨링의 높은 난이도로 인해, 데이터셋 확보 어려움
- **Contribution / Key idea**
  - Active Label Correction (ALC) framework
    - Correction query
    - Look-ahead acquisition function

전유진

Active Label Correction Framework

- **Overview**

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)
8: **end for**
9: **return** $\mathcal{D}_T$ and θ_T

Kim, Hoyoung, et al. "Active label correction for semantic segmentation with foundation models." *arXiv preprint arXiv:2403.10820* (2024).

Active Label Correction Framework

- **Initial Dataset Preparation**

❑ Grounded SAM을 통해 초기 레이블 생성

(a) Unlabeled image (b) Grounded-SAM

❑ 모델 학습

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

전유진

Kim, Hoyoung, et al. "Active label correction for semantic segmentation with foundation models." *arXiv preprint arXiv:2403.10820* (2024).

Active Label Correction Framework

- **Correction Query**

  - 기존 Classification query과 달리 Correction query는 pseudo-label이 맞다면 패스할수 있어서 비용 절약

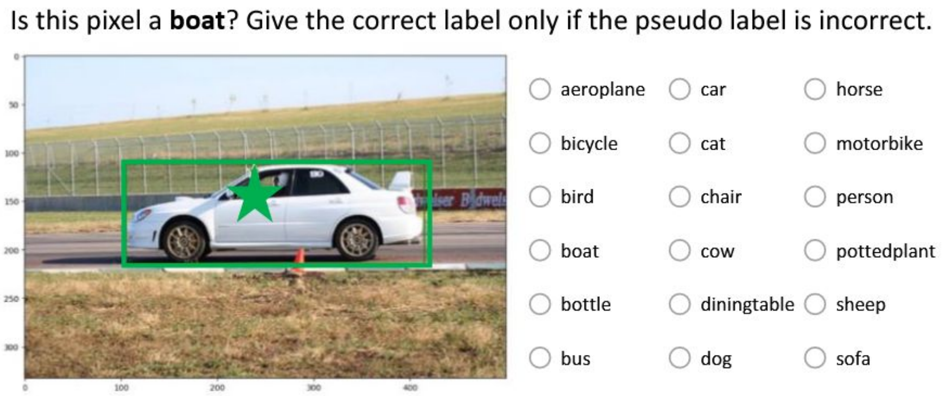

Figure 2: *An example of correction query*. Correction query presents an instruction requesting a label for a representative pixel (green star), an image displaying an object within a bounding box (green rectangle), and possible class options.

전유진

Kim, Hoyoung, et al. "Active label correction for semantic segmentation with foundation models." *arXiv preprint arXiv:2403.10820* (2024).

Active Label Correction Framework

- **Diversified Pixel Pool**

  - **대표픽셀 집합 생성**

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

  - 각 슈퍼픽셀 내에서 dominant 영역과 유사한 픽셀을 대표픽셀로 설정

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

  - 슈퍼픽셀 s의 dominant 레이블

$$\mathrm{D}_\theta(s) := \arg\max_{c \in \mathcal{C}} |\{x \in s : y_\theta(x) = c\}|, \quad (5)$$

  - 슈퍼픽셀 s 내에서 dominant 영역

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

전유진

Kim, Hoyoung, et al. "Active label correction for semantic segmentation with foundation models." *arXiv preprint arXiv:2403.10820* (2024).

Active Label Correction Framework

- **Look-Ahead Acquisition Function**

❑ 쿼리할 B개의 픽셀 샘플링 from 대표픽셀 집합

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

❑ 픽셀 x의 불확실성 (**CIL**)

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

❑ **최종 획득함수 (SIM): 슈퍼픽셀 s 내의 픽셀들의 CIL 값들을 대표픽셀과의 유사도 기반 가중 평균한 값**

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

❑ 픽셀 x의 불확실성 (**LCIL**): 슈퍼픽셀 s 내의 픽셀들의 CIL 값들을 균등 가중 평균한 값

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

❑ 업데이트된 데이터셋으로 모델 학습하며 t 라운드 종료

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

전유진

Kim, Hoyoung, et al. "Active label correction for semantic segmentation with foundation models." *arXiv preprint arXiv:2403.10820* (2024).