[Active Label Correction Using Robust Parameter Update and Entropy Propagation]

ECCV 2022

Kwang In Kim UNIST, Korea

- Problem / objective
  - ➔ 기존 ALC* 접근법들의 *entropy 가 높은 샘플을 쿼리하는 방식 의 부정확함*

- Contribution / Key idea
  - ➔ 새로운 ALC 알고리즘 제안 : *robustly estimate the entropy*

ALC* = Active Label Correction

전유진

Active Label Correction 이란?

1. Labeled dataset 으로 모델 학습
2. 모델이 제일 중요하다고 생각하는 noisy label 일것 같은 샘플 제안
3. Annotator 가 제안받은 샘플의 label 수정
4. 수정된 labeled dataset 으로 모델 학습
5. 2-4 과정을 할당된 cost 다쓸때까지 또는 성능 만족할때까지 반복

Kwang In Kim UNIST, Korea

Active Label Correction 이란?

1. Labeled dataset 으로 모델 학습
2. *모델이 제일 중요하다고 생각하는 noisy label 일것 같은 샘플 제안* ← *보통 entropy 가 높은 샘플 제안*
3. Annotator 가 제안받은 샘플의 label 수정
4. 수정된 labeled dataset 으로 모델 학습
5. 2-4 과정을 할당된 cost 다쓸때까지 또는 성능 만족할때까지 반복

Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim UNIST, Korea

ECCV 2022

기존 ALC 접근법들의 'high entropy 샘플 쿼리 방식' 의 문제

1. Unreliable Confidence Estimates

   모델이 noisy dataset 에 이미 학습이 되어버려서, 이를 사용하여 측정한 entropy 를 믿을 수 없음.

2. Redundant Labeling

   무작정 entropy 높은 샘플들만 고르다보면 불필요한 쿼리 초래함

➔ **Ours : 'high entropy 샘플 쿼리' 하긴 할건데, robustly estimate the entropy.**

1. Robust Parameter Update
2. Entropy Propagation

전유진

Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim UNIST, Korea

ECCV 2022

Algorithm

Algorithm 1 Robust active label correction algorithm

Input: Noisy data $D = \{(x_i, y_i)\}_{i=1}^N$, and labeling batch size Q and budget G.
Initialization: $D^0 = \emptyset$, $B^0 = \emptyset$, $S^0 = \emptyset$, and $t = 1$.
repeat
  $\alpha_j = \frac{1}{N}$ for $j = 1 \ldots, N$.
  if $\mod(t, Q) = 0$ **then**
    $S^t = S^{t-1} \cup B^{t-1}$.
    repeat
      Update the classifier h parameter W according to Eq 2.
      Update the contribution coefficients $\{\alpha_j\}_{j=1}^N$ using Eq 3.
    until maximum epoch reached.
    Train h with $\{\alpha_j\}_{j=1}^N$.
    Evaluate the entropy values $\mathbf{e}$ using h on D.
    $B^t = \emptyset$.
  end if
  Sample the candidate set C from $D \setminus S^t$ using p^S.
  Select an example $x_* \in C$ with the largest entropy e.
  repeat
    Assign zero to $e(x_*)$.
    Update $\mathbf{e}$ using Eq. 8.
  until maximum diffusion steps reached.
  $B^t = B^t \cup \{x_*\}$.
  $t = t + 1$.
until labeling budget reached.
Output: Trained classifier h^t and (partially) cleaned label set S^t.

전유진

Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim UNIST, Korea

ECCV 2022

Robust Parameter Update

1. 매 epoch 마다 global weights update

$$\{\alpha_j\}_{j=1}^N \qquad \alpha_j \geq 0, \sum_{j=1}^N \alpha_j = 1 \qquad \widehat{D} = \{(x_j, y_j)\}_{j=1}^N$$

$\boldsymbol{\alpha}(1) = [\alpha_1(1), \ldots, \alpha_N(1)]^\top$: 1 에포크일때의 global weights : uniformly initialized

2. 매 iter 의 contribution parameters $\{\overline{\alpha}_k(i)\}$ 는 global weights $\{\alpha_j\}_{j=1}^N$ 로부터 선택

3. 매 iter 마다 parameter update

$$W(i+1) = W(i) - \eta(i) \sum_{(x_k, y_k) \in D_i} \overline{\alpha}_k(i) \nabla_W l(h(x_k), y_k), \qquad (2)$$

처음에는 clean, noisy data 모두 파라미터 업데이트에 동등하게 기여.

Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim UNIST, Korea

ECCV 2022

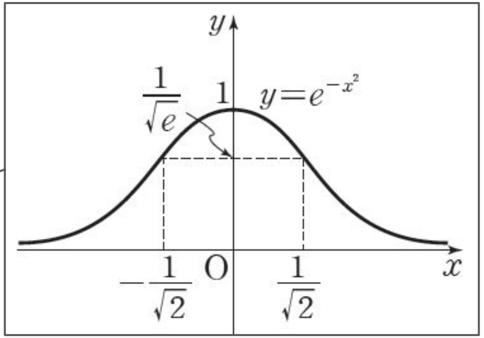
Robust Parameter Update

1. 매 epoch 마다 global weights update

$$\boldsymbol{\alpha}(q+1) = (1-\delta^\alpha)\boldsymbol{\alpha}(q) + \delta^\alpha \frac{\mathbf{g}(q)}{\|\mathbf{g}(q)\|_1}, \qquad (3)$$

$$\mathbf{g}(q) = [g(l(h^q(x_1), y_1)), \ldots, g(l(h^q(x_N), y_N))]^\top$$

$$g(z) = \exp\left(-\frac{z^2}{\sigma^\alpha}\right) \qquad (4)$$



2. 매 iter 의 contribution parameters $\{\overline{\alpha}_k(i)\}$ 는 global weights $\{\alpha_j\}_{j=1}^N$ 로부터 선택

3. 매 iter 마다 parameter update

$$W(i+1) = W(i) - \eta(i) \sum_{(x_k, y_k)\in D_i} \overline{\alpha}_k(i) \nabla_W l(h(x_k), y_k), \qquad (2)$$

파라미터 업데이트에 noisy data 의 영향 줄임

전유진

Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim UNIST, Korea

ECCV 2022

Entropy Propagation

1. query 한 데이터의 엔트로피를 0으로 설정
2. 공간적으로 query 한 데이터 주변에 있는 데이터들에게 업데이트된 엔트로피 전파

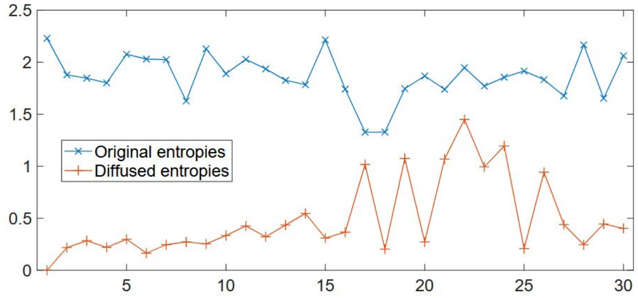


Fig. 1: An example of entropy diffusion on *CIFAR-10* dataset. A point x_* is newly labeled and the corresponding entropy is updated to zero. The entropies of the remaining examples in D are accordingly adjusted. The x-axis shows the indices of data points ordered inversely according to the distance to x_*. The first entry is x_*. When x_* has originally a high entropy value, its spatial neighbors also exhibit high entropy values (the average entropy on D was less than 0.7). Applying diffusion on D suppressed the entropies of points near x_*. Note that the degrees of suppression are proportional to the similarity to x_*.

전유진

Entropy Propagation

1. Diffusion of smooth function $g \in C^{\infty}(\mathcal{X})$

$$\frac{\partial g}{\partial t} = \Delta_p g, \quad (5)$$

: 위치 x 에 있는 질량 g(x) 를 p 로 가중하여 전체 manifold $\mathcal{X}$ 로 점진적으로 전파하는 과정

2. Diffusion of entropy values $\mathbf{e} = [e(x_1), \ldots, e(x_N)]^\top$

$X = \{x_j\}_{j=1}^N$: embedding of $\mathcal{X}$ onto a Euclidean space

$$\frac{\partial \mathbf{e}}{\partial t} = -L\mathbf{e}, \quad (6)$$

3. Entropy propagation algorithm

$$\mathbf{e}(i+1) = \mathbf{e}(i) - \delta^L L\mathbf{e}(i) \quad (8)$$

time-discretization step size

Active Label Correction U

Kwang In Kim UNIST, Kore

ECCV 2022

Experiments

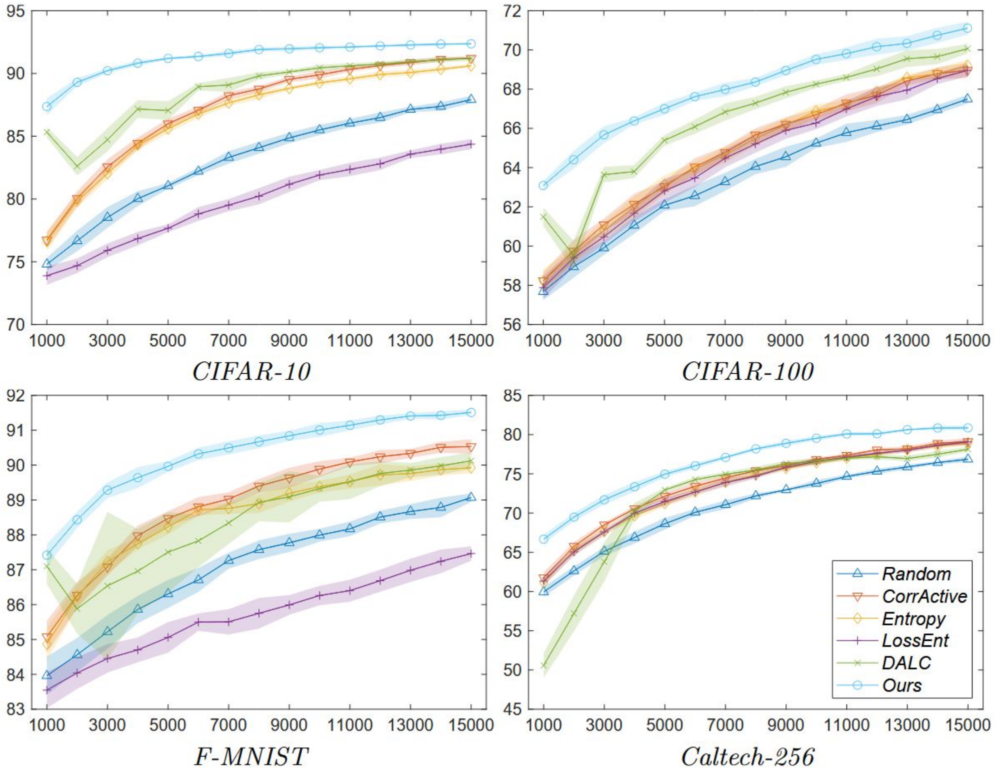


Fig. 2: Mean accuracy (%) with standard deviation (shaded) of different active label correction algorithms under uniform noise. The x-axis corresponds to the number of queried labels. All ALC algorithms outperformed *Random* except for *LossEnt* on *CIFAR-10* and *F-MNIST*. *DALC* demonstrated competitive performance in *CIFAR-10*, *CIFAR-100*, and later learning stages of *Caltech-256*. Our algorithm achieved further significant and consistent improvements.

전유진

Active Label Correction Using Robust Parameter Update and Entropy Propagation

Kwang In Kim UNIST, Korea

ECCV 2022

Experiments

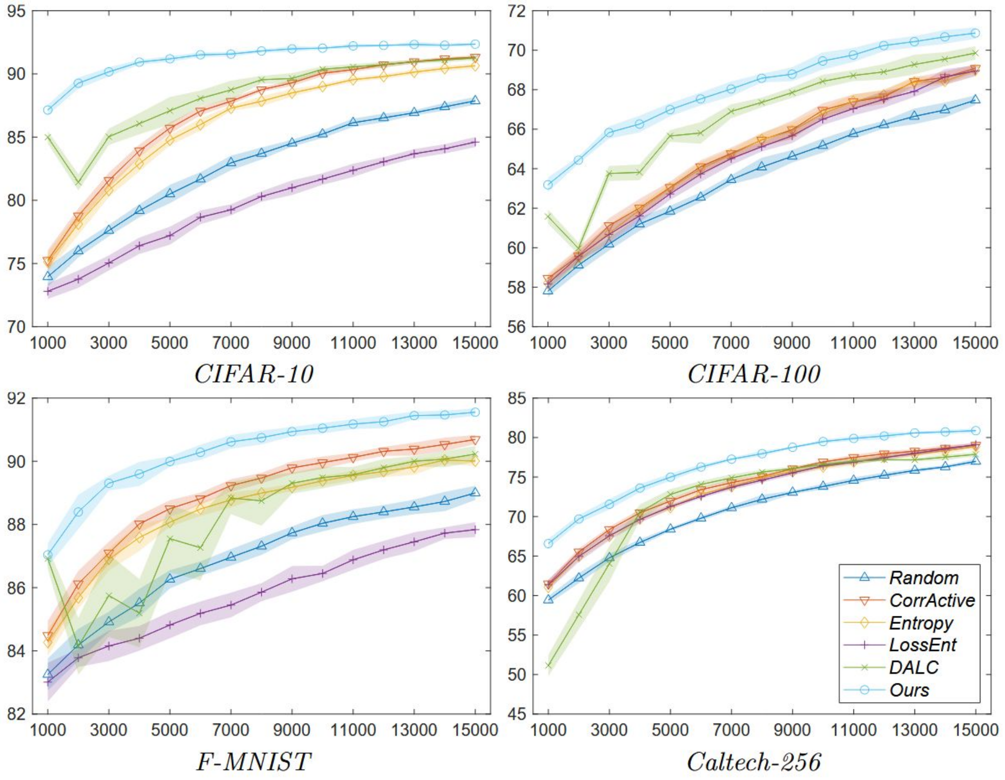


Fig. 3: Mean accuracy (%) with standard deviation (shaded) of different active label correction algorithms under class-symmetry flipping noise.

전유진

Experiments

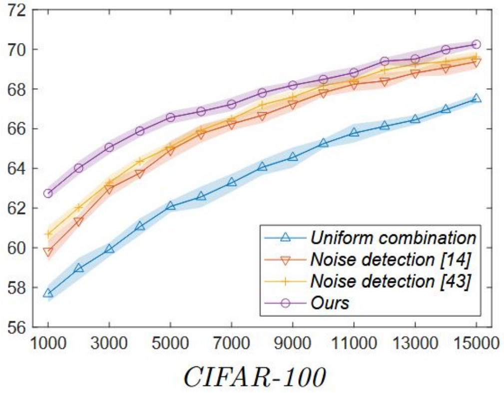


Fig. 4: Performance of our robust parameter update approach (Eq. 2), the standard uniform gradient combination, and the explicit noise detection methods of [14,43]. x- and y-axes show the number of acquired labels and the corresponding classification accuracy (%), respectively. Our *soft* gradient combination approach provides considerably higher labeling efficiency than *hard* noise detection and uniform gradient averaging.

Experiments

Table 1: Average noisy example selection accuracy (%) of different ALC algorithms defined as the ratio between the number of queried points in $D \setminus D^C$ and the total number of queries; *CIFAR-100*. Our algorithm consistently achieved the highest selection accuracy.

# labels	1,000	3,000	5,000	7,000	9,000	11,000	13,000	15,000
Random	81.20	80.50	81.50	82.00	80.50	80.00	81.00	82.10
Entropy	90.30	91.30	90.00	91.50	92.60	90.70	90.80	91.40
LossEnt	93.70	93.98	93.02	93.41	93.70	93.79	94.27	93.22
CorrActive	87.60	89.70	87.50	87.60	89.20	89.10	89.70	90.20
DALC	89.10	92.10	91.70	91.10	91.10	90.20	88.30	88.80
Ours	97.90	98.00	97.30	97.80	97.80	97.80	98.10	97.60