Estimating label quality and errors in semantic segmentation data via any model

Vedang Lad 1 2 Jonas Mueller 1

- **Problem/Objective**
  - Noisy label
- **Contribution/Key Idea**
  - Noisy label detection
    - 라벨 품질 평가하는 Softmin 알고리즘 제안
    - 라벨 품질이 낮다는 것은 부정확한 라벨이 많다는 것

Estimating label quality and errors in semantic segmentation data via any model

Vedang Lad 1 2 Jonas Mueller 1

- **Correctly Classified Pixels (CCP)**

  모델 예측 결과와 라벨이 일치하는 비율

$$s_{CCP}(x) = \frac{\sum_{i,j} \mathbb{I}[l_{ij} = P_{ij}]}{h \cdot w} \tag{1}$$

- **Thresholded CCP (TCCP)**

  CCP 에 클래스 별 임계값 도입

$$s^k_{TCCP,t}(x) = \frac{\sum_{i,j} \mathbb{I}[l_{ij} = k, p_{ijk} > \tau]}{h \cdot w} \tag{2}$$

$$\tau^*_k = \underset{\tau \in T}{\operatorname{argmax}}\, s^k_{TCCP,\tau}(x) \tag{3}$$

$$s_{TCCP}(x) = \frac{1}{K} \sum_{k=1}^{K} s^k_{TCCP,\tau^*_k}(x) \tag{4}$$

Estimating label quality and errors in semantic segmentation data via any model

ICML 2023 Workshop

Vedang Lad 1 2 Jonas Mueller 1

- **Confidence In Label (CIL)**

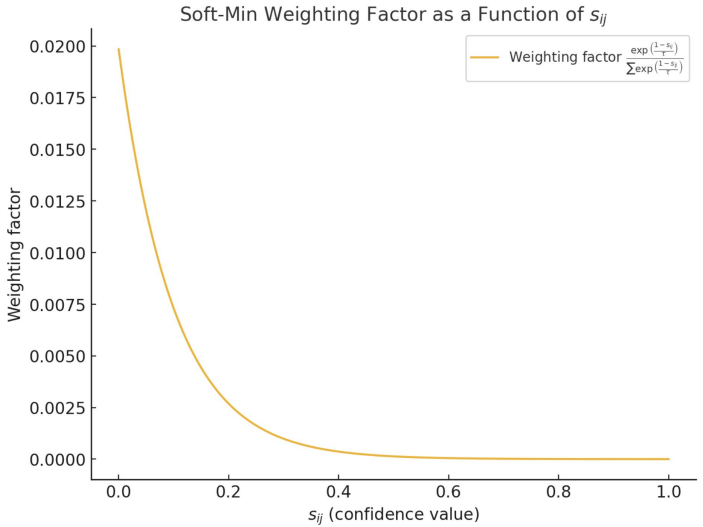
  픽셀마다 라벨 품질 점수 구하고 이미지 단위 평균

$$s_{CIL}(x) = \frac{1}{h \cdot w} \sum_{i,j} s_{ij} \qquad (5)$$
$$s_{ij} := p_{i,j,l_{ij}}$$

- **Softmin (Our Proposed Method)**

  라벨 품질 최소 점수를 soft approximation

$$s_{SM}(x) = \sum_{i,j} s_{ij} \cdot \frac{\exp\left(\frac{1-s_{ij}}{\tau}\right)}{\sum_{i,j} \exp\left(\frac{1-s_{ij}}{\tau}\right)} \qquad (6)$$



Estimating label quality and errors in semantic segmentation data via any model

Vedang Lad 1 2 Jonas Mueller 1

- **Confident Learning Counts (CLC)**

  이미지 내 라벨링 잘된 픽셀 비율

$$s_{CLC} = \frac{\sum_{i,j} b_{ij}}{h \cdot w} \qquad (7)$$

- **IOU**

  예측 마스크와 레이블 마스크 간 IoU

$$s_{IOU}(x) = \frac{|\mathbf{P} \cap \mathbf{l}|}{|\mathbf{P} \cup \mathbf{l}|} \qquad (8)$$

- **Connected Components (CoCo)**

  각 이미지 내 모든 연결 성분들의 평균 라벨 품질 점수 $s_{CoCo}(x)$

Estimating label quality and errors in semantic segmentation data via any model

ICML 2023 Workshop

Vedang Lad 1 2 Jonas Mueller 1

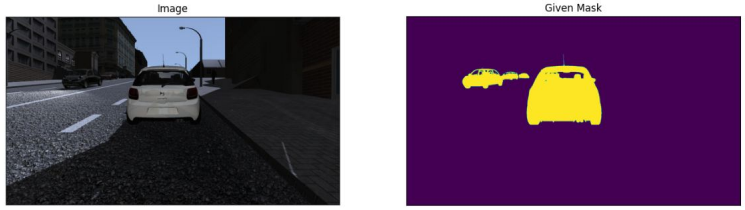
- **Experiments**

라벨 오류 유형 3가지

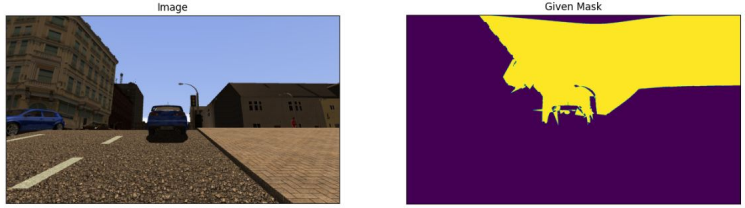
학습 및 검증 이미지 개수 : 1112개, 1112개

검증 세트에 대해서만 예측 및 라벨 품질 점수 계산함.

Drop (Car overlooked) - Given Mask for Unlabeled class



Shift Error (in Sky class) - Given Mask for Sky class



Swap (Building → Vegetation) - Given Mask for Vegetation

Estimating label quality and errors in semantic segmentation data via any model

ICML 2023 Workshop

Vedang Lad 1 2 Jonas Mueller 1

- **Experiments**

Table 1. AUROC achieved by various label quality scores used with two types of models for detecting three types of annotation errors.

METHOD	DROP		SWAP		SHIFT	
MODEL	DEEPLABV3+	FPN	DEEPLABV3+	FPN	DEEPLABV3+	FPN
CORRECTLY CLASSIFIED PIXELS	0.915	0.916	**1.000**	**1.000**	0.833	0.798
THRESHOLDED CCP	0.869	0.869	0.993	0.993	0.852	0.818
CONFIDENCE IN LABEL	0.915	0.916	**1.000**	**1.000**	0.832	0.797
CONFIDENT LEARNING COUNTS	0.904	0.905	**0.999**	**0.999**	0.807	0.778
IOU	0.921	0.901	0.880	0.880	0.713	0.649
CONNECTED COMPONENTS	0.880	0.888	0.984	0.982	0.783	0.819
SOFTMIN	**0.951**	**0.947**	**0.998**	**0.998**	**0.863**	**0.828**

Estimating label quality and errors in semantic segmentation data via any model

ICML 2023 Workshop

Vedang Lad 1 2 Jonas Mueller 1

- **Experiments**

Table 2. AUPRC achieved by various label quality scores used with two types of models for detecting three types of annotation errors.

Error Type	Drop		Swap		Shift	
Model	DeepLabV3+	FPN	DeepLabV3+	FPN	DeepLabV3+	FPN
Correctly Classified Pixels	0.814	0.814	**0.999**	**0.999**	0.474	0.399
Thresholded CCP	0.684	0.681	0.983	0.983	0.526	0.440
Confidence In Label	0.813	0.814	**0.999**	**0.999**	0.472	0.398
Confident Learning Counts	0.795	0.796	**0.999**	**0.999**	0.429	0.375
IoU	0.808	0.749	0.754	0.759	0.440	0.320
Connected Components	0.654	0.675	0.971	0.965	0.519	**0.537**
Softmin	**0.888**	**0.875**	**0.996**	**0.996**	**0.545**	0.461