DMQA Open Seminar

Label Noise Learning with In- and Out-of-Distribution Noise

2026.03.06

고려대학교 산업경영공학과

Data **M**ining & **Q**uality **A**nalytics Lab.

이정민

KOREA UNIVERSITY

DMQA

발표자 소개

❖ **이정민(JungMin Lee)**

- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab.(김성범 교수님)
- 석박 통합 과정(2022.03~Present)

❖ **Research Interest**

- Uncertainty Quantification
- Label Noise Learning
- Vision-Language Models

❖ **Contact**

- jungmin9195@korea.ac.kr

Contents

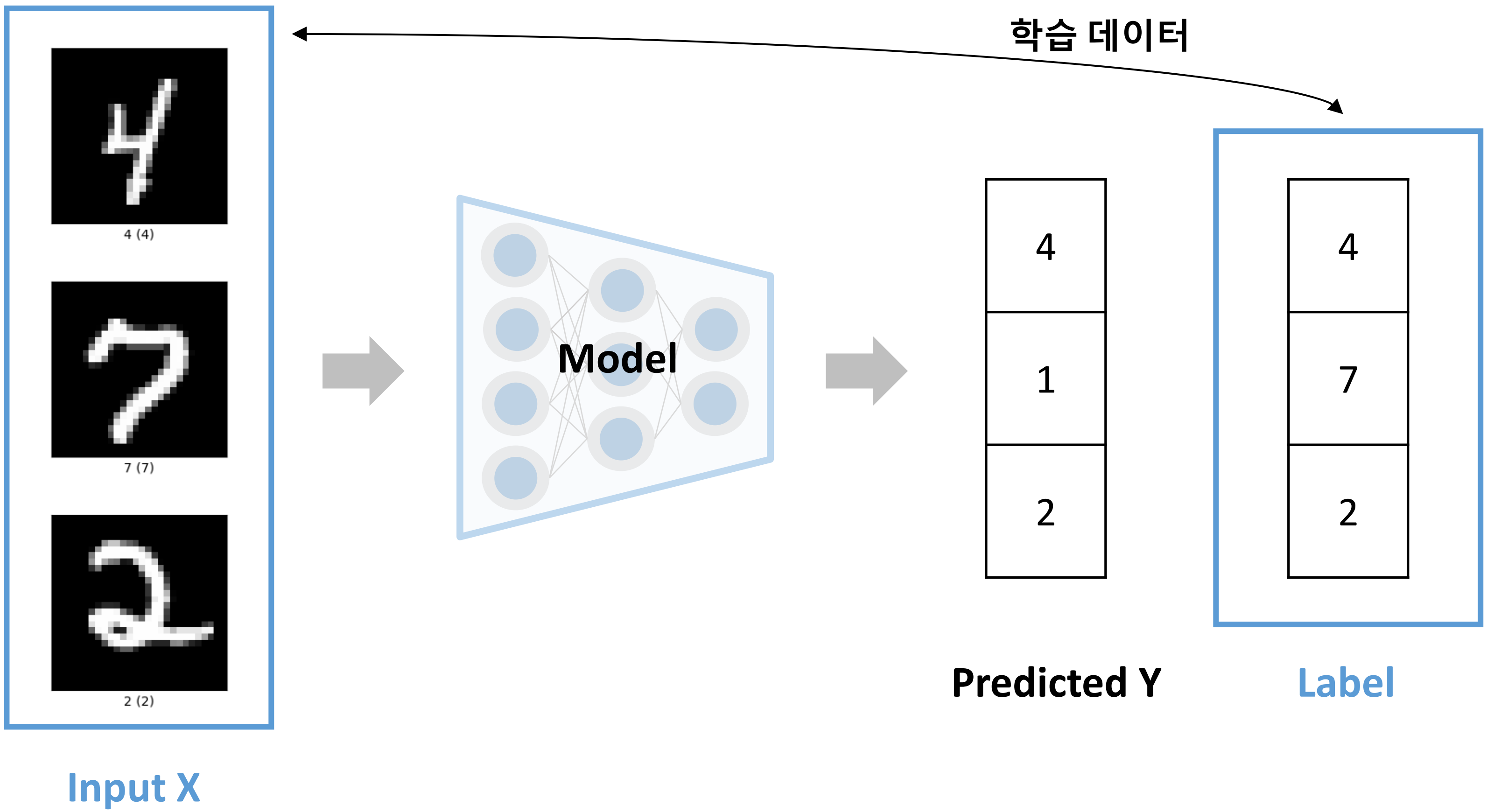
Introduction

Introduction

Noisy Label

❖ Supervised Learning

- 일반적인 supervised Learning에서는 학습 데이터의 label이 모두 **정확함**을 가정
- 정확한 label을 통해 입력 데이터와 label의 관계를 모델이 학습



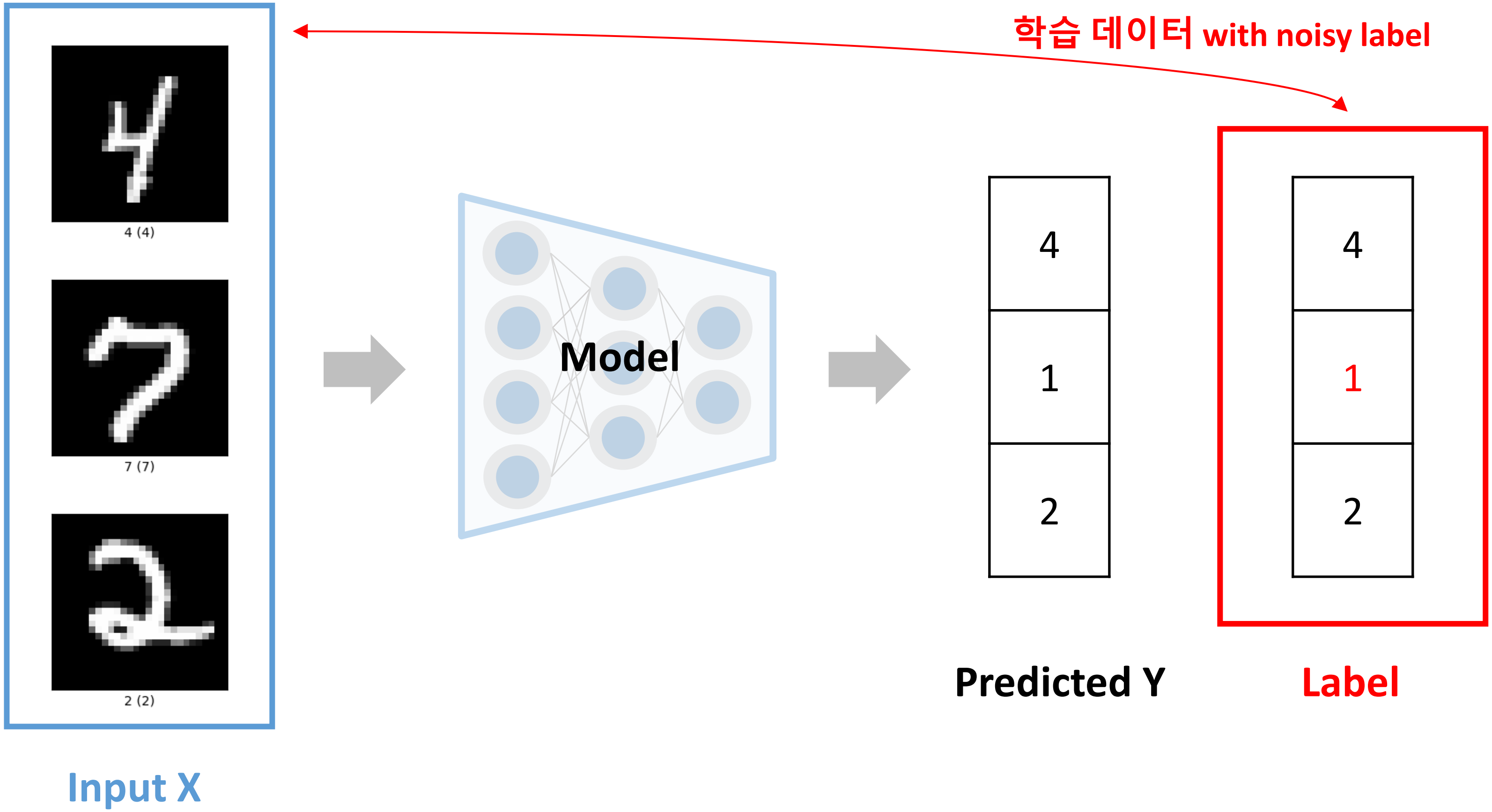
DMQA

Introduction

Noisy Label

❖ Noisy Labels

- 현실에는 label이 정제되지 않은 방대한 양의 데이터가 존재
- Noisy label로 인해 **모델의 일반화 성능이 저하됨**



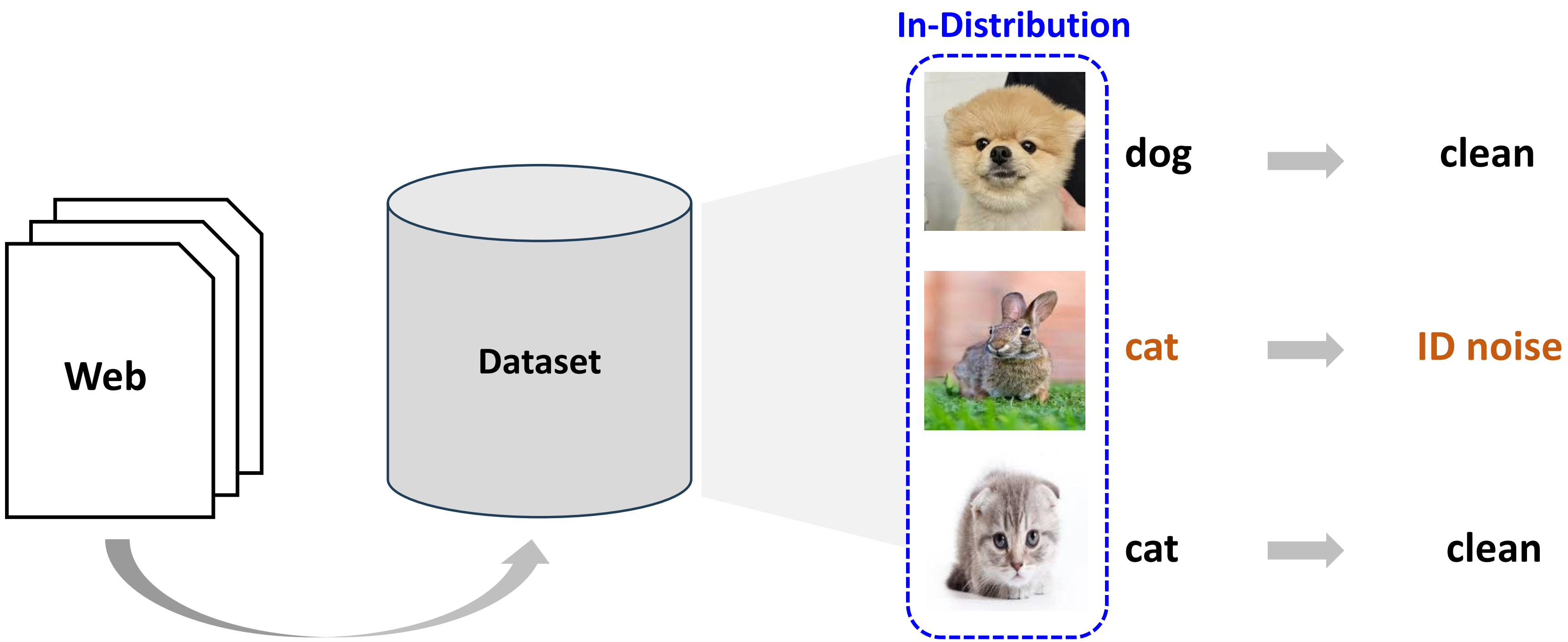
DMQA

Introduction

Noisy Label

❖ In-Distribution (ID) Noise

- 수집하고자 하는 데이터셋 내에서 잘못된 labeling 발생



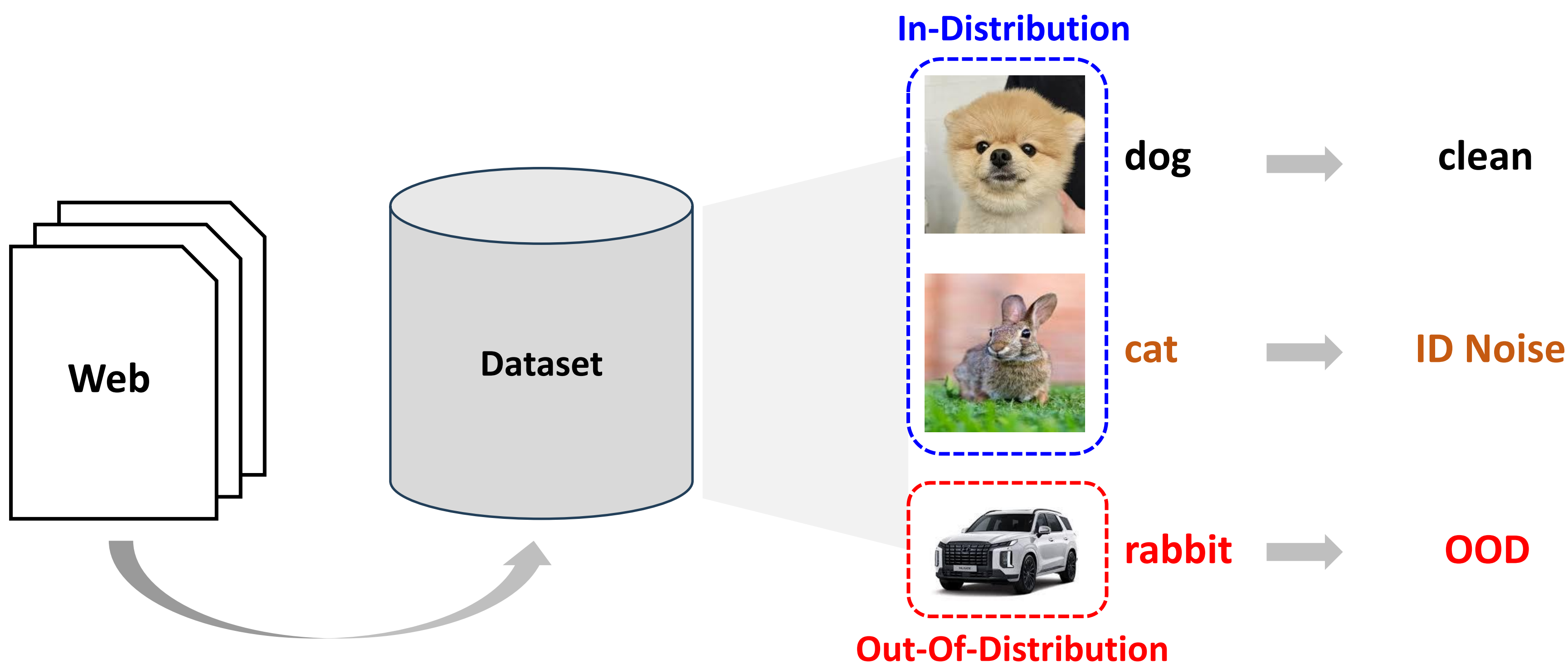
DMQA

Introduction

Noisy Label

❖ In- and Out-of-Distribution (ID & OOD) Noise

- Web을 통해 이미지 데이터셋을 수집하다 보면, **분포가 아예 다른 종류의 데이터**가 같이 수집될 수 있음
- 학습 데이터에 out-of-distribution(OOD)와 in-distribution(ID) noise가 혼재함을 가정



Introduction

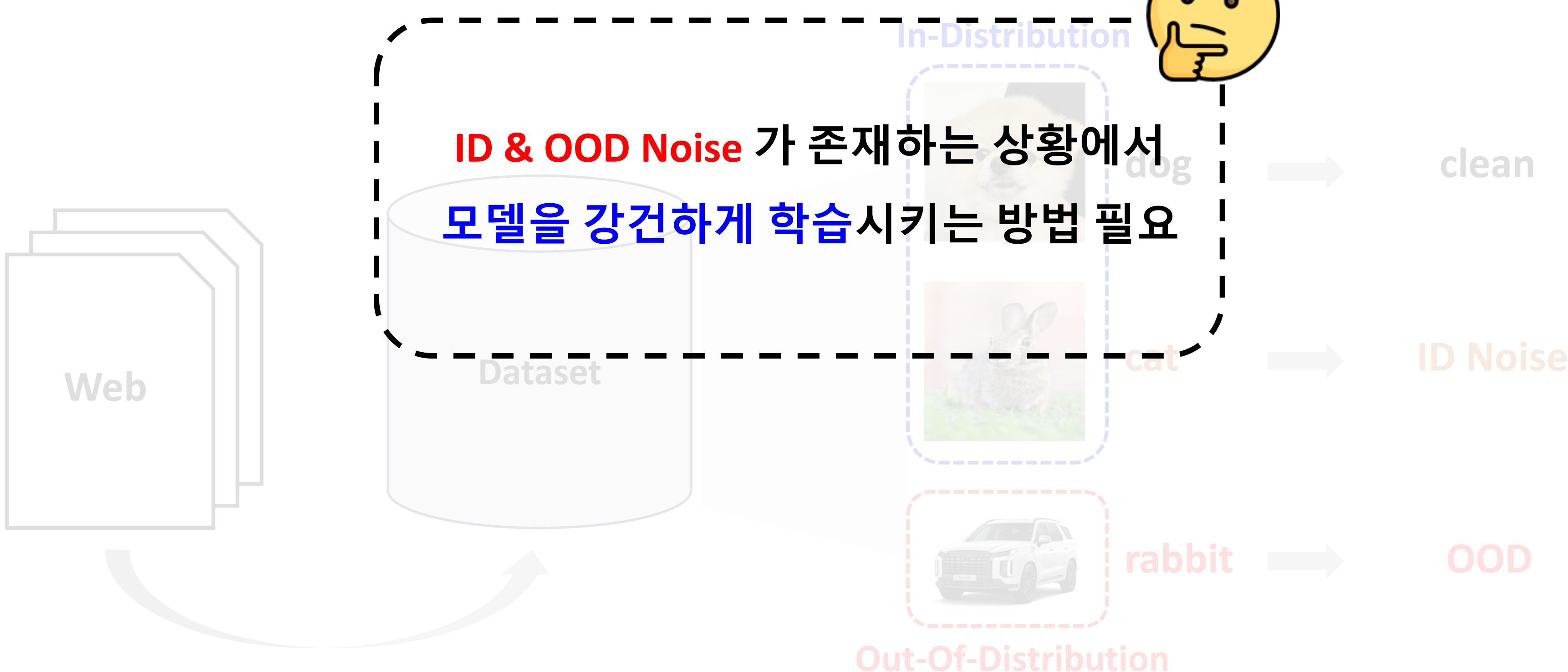
Noisy Label

❖ In- and Out-of-Distribution (ID & OOD) Noise

- Web을 통해 이미지 데이터셋을 수집하다 보면, 분포가 아예 다른 종류의 데이터가 같이 수집될 수 있음
- 학습 데이터에 out-of-distribution(OOD)와 in-distribution(ID) noise가 혼재함을 가정

ID & OOD Noise 가 존재하는 상황에서

모델을 강건하게 학습시키는 방법 필요



Introduction

Related Seminar

❖ Label Noise Learning

종료

Deep Neural Networks with Noisy Labels

김상훈
Data Mining & Quality Analytics Lab.
2022.08.28

Deep Neural Networks with Noisy Labels

발표자: 김상훈

2022년 8월 28일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료

Noisy Label Learning

김상훈
Data Mining & Quality Analytics Lab.

Noisy Label Learning

발표자: 김상훈

2023년 6월 3일

오전 12시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

종료

Beyond DivideMix:
Advances in Label Noise Learning
With Semi-Supervised Learning

2025.08.01
고려대학교 산업경영공학과
Data Mining & Quality Analytics Lab.

Beyond DivideMix: Advances in Label Noi

발표자: 이정민

2025년 8월 1일

오전 12시 ~

고려대학교 신공학관 218호

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

DMQA

Label Noise Learning with In- and Out-of-Distribution Noise

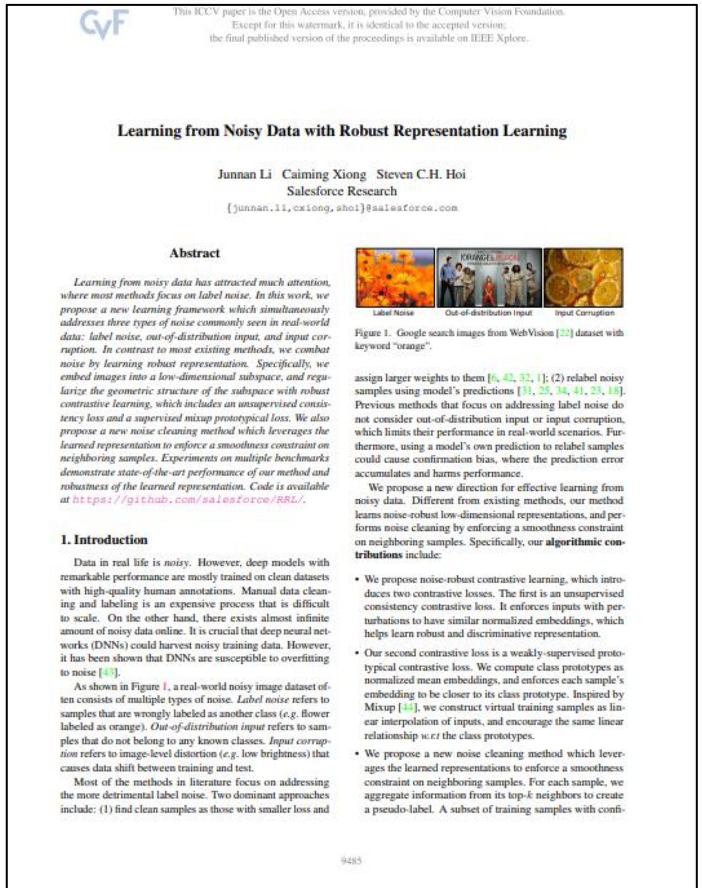
DMQA

RRL

Paper

❖ Learning from Noisy Data with Robust Representation Learning (2021, ICCV)

- **OOD까지 결합된** label noise learning 문제 상황을 처음으로 제시



Learning from Noisy Data with Robust Representation Learning

Junnan Li Caiming Xiong Steven C.H. Hoi
Salesforce Research
{junnan.li,cxiong,shoi}@salesforce.com

Abstract

Learning from noisy data has attracted much attention, where most methods focus on label noise. In this work, we propose a new learning framework which simultaneously addresses three types of noise commonly seen in real-world data: label noise, out-of-distribution input, and input corruption. In contrast to most existing methods, we combat noise by learning robust representation. Specifically, we embed images into a low-dimensional subspace, and regularize the geometric structure of the subspace with robust contrastive learning, which includes an unsupervised consistency loss and a supervised mixup prototypical loss. We also propose a new noise cleaning method which leverages the learned representation to enforce a smoothness constraint on neighboring samples. Experiments on multiple benchmarks demonstrate state-of-the-art performance of our method and robustness of the learned representation. Code is available at https://github.com/salesforce/RRL/.

1. Introduction

Data in real life is *noisy*. However, deep models with remarkable performance are mostly trained on clean datasets with high-quality human annotations. Manual data cleaning and labeling is an expensive process that is difficult to scale. On the other hand, there exists almost infinite amount of noisy data online. It is crucial that deep neural networks (DNNs) could harvest noisy training data. However, it has been shown that DNNs are susceptible to overfitting to noise [43].

As shown in Figure 1, a real-world noisy image dataset often consists of multiple types of noise. *Label noise* refers to samples that are wrongly labeled as another class (*e.g.* flower labeled as orange). *Out-of-distribution input* refers to samples that do not belong to any known classes. *Input corruption* refers to image-level distortion (*e.g.* low brightness) that causes data shift between training and test.

Most of the methods in literature focus on addressing the more detrimental label noise. Two dominant approaches include: (1) find clean samples as those with smaller loss and

Figure 1. Google search images from WebVision [22] dataset with keyword "orange".

assign larger weights to them [6, 42, 32, 1]; (2) relabel noisy samples using model's predictions [31, 25, 34, 41, 23, 18]. Previous methods that focus on addressing label noise do not consider out-of-distribution input or input corruption, which limits their performance in real-world scenarios. Furthermore, using a model's own prediction to relabel samples could cause confirmation bias, where the prediction error accumulates and harms performance.

We propose a new direction for effective learning from noisy data. Different from existing methods, our method learns noise-robust low-dimensional representations, and performs noise cleaning by enforcing a smoothness constraint on neighboring samples. Specifically, our **algorithmic contributions** include:

- We propose noise-robust contrastive learning, which introduces two contrastive losses. The first is an unsupervised consistency contrastive loss. It enforces inputs with perturbations to have similar normalized embeddings, which helps learn robust and discriminative representation.
- Our second contrastive loss is a weakly-supervised prototypical contrastive loss. We compute class prototypes as normalized mean embeddings, and enforces each sample's embedding to be closer to its class prototype. Inspired by Mixup [44], we construct virtual training samples as linear interpolation of inputs, and encourage the same linear relationship *w.r.t* the class prototypes.
- We propose a new noise cleaning method which leverages the learned representations to enforce a smoothness constraint on neighboring samples. For each sample, we aggregate information from its top-k neighbors to create a pseudo-label. A subset of training samples with confi-

9485

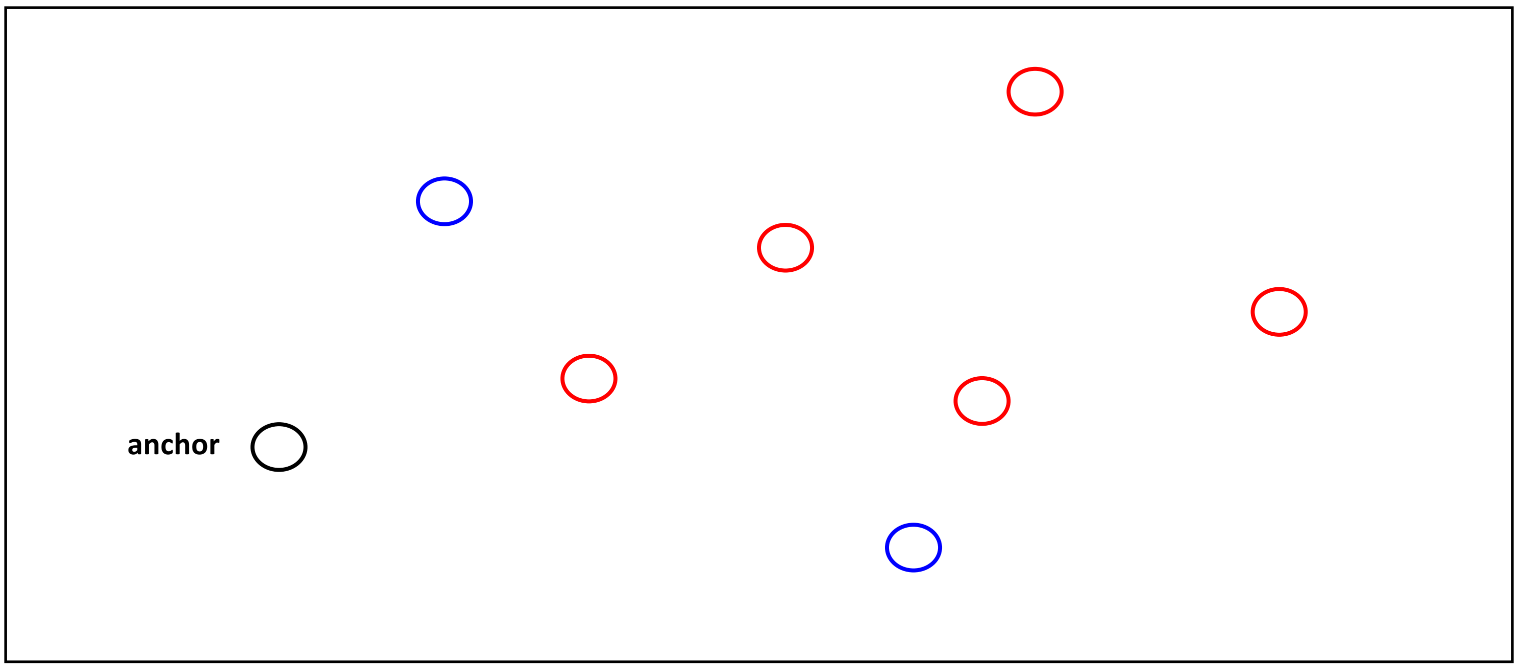
Li, J., Xiong, C., & Hoi, S. C. (2021). Learning from noisy data with robust representation learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9485-9494).

RRL

Background

❖ Contrastive learning

- Metric learning 방법론 중 하나로, 데이터 간 유사도 정보를 통해 거리 함수를 학습하는 방법론
- Main idea: Anchor를 기준으로 positive samples는 가깝도록, negative samples는 멀도록 학습



[Embedding space]

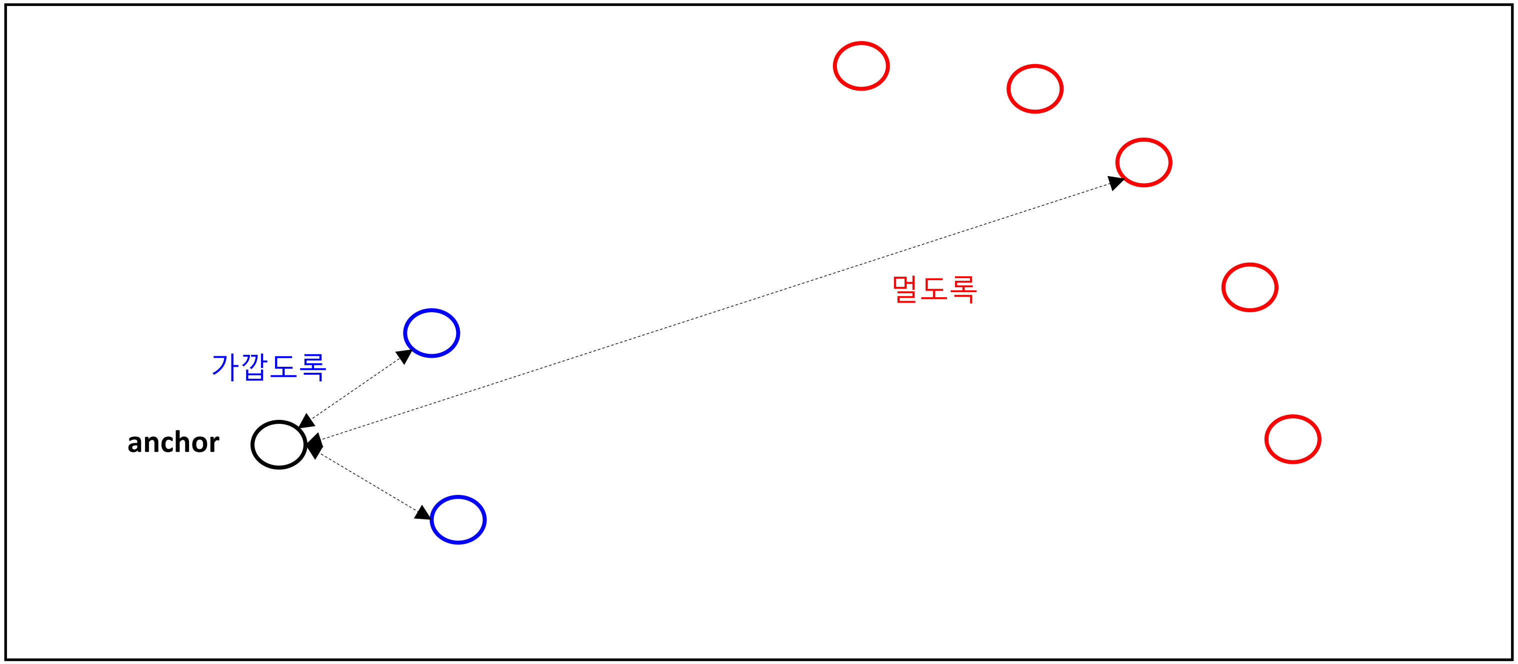
DMQA

RRL

Background

❖ Contrastive learning

- Metric learning 방법론 중 하나로, 데이터 간 유사도 정보를 통해 거리 함수를 학습하는 방법론
- Main idea: Anchor를 기준으로 positive samples는 가깝도록, negative samples는 멀도록 학습



[Embedding space]

DMQA

RRL

Background

❖ 문제 상황

- Label noise + OOD + Input corruption

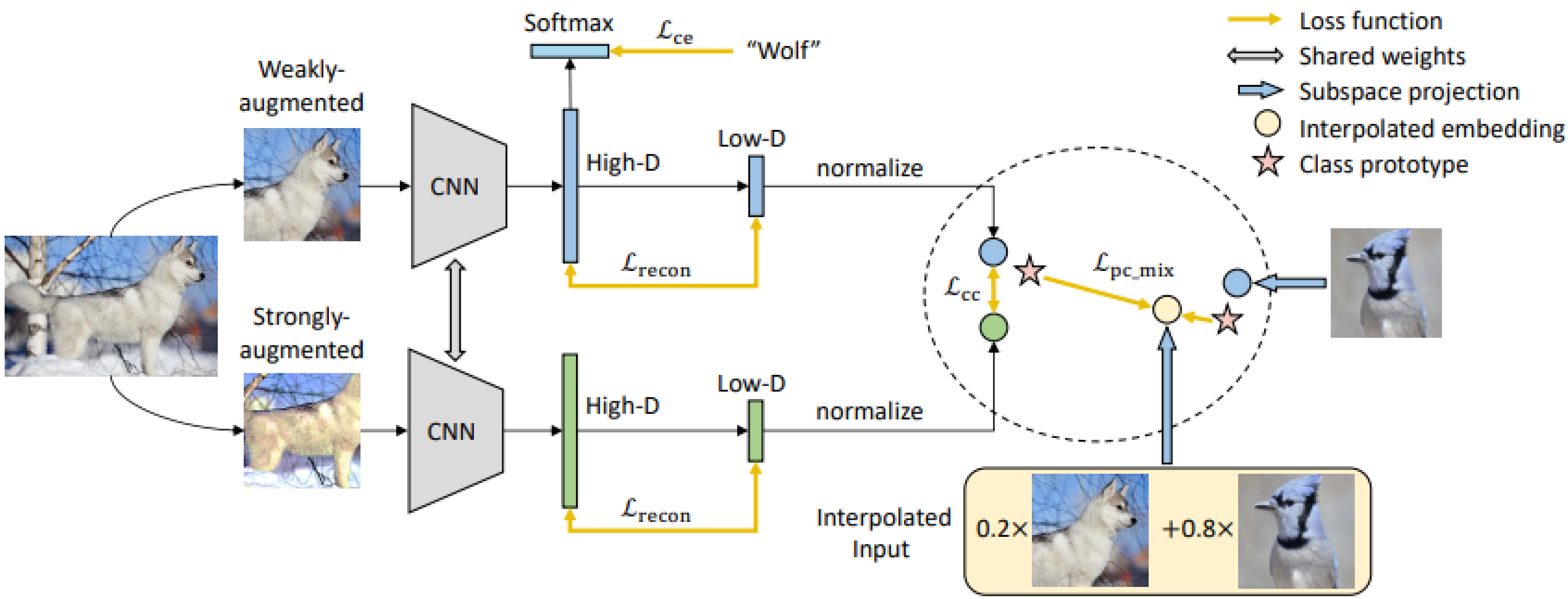
Figure 1. Google search images from WebVision [22] dataset with keyword "orange".

RRL

Method

❖ Overall Process

- 학습 대상: Encoder + Classifier + Linear autoencoder

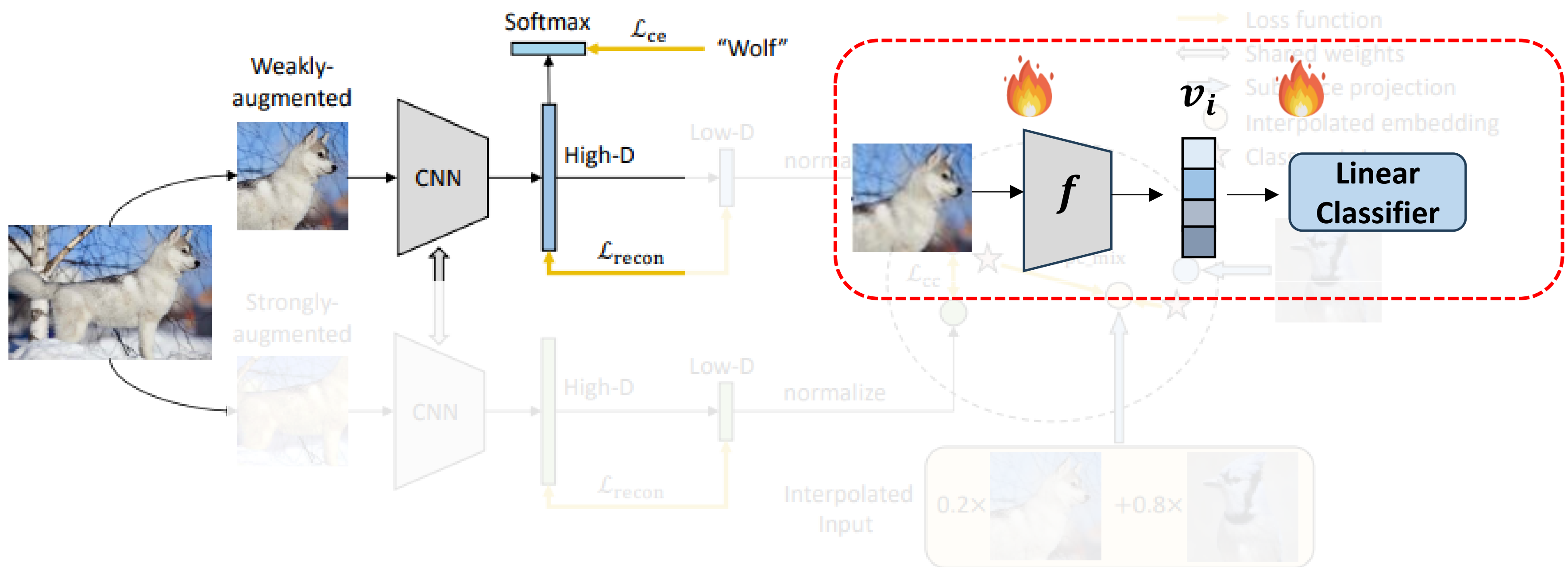


DMQA

RRL

Method

1. Cross-entropy loss $(L_{ce}) = -\sum_{i=1}^{b} \log p(y_i; x_i)$



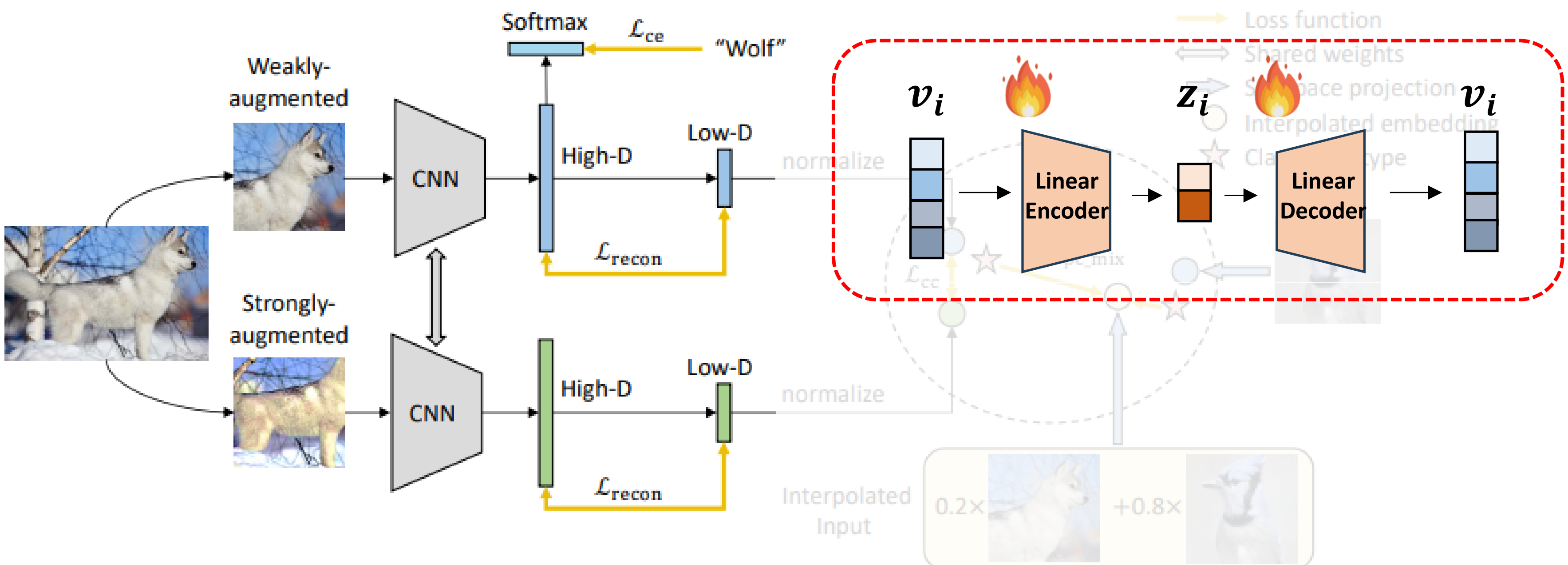
DMQA

RRL

Method

2. Reconstruction loss $(L_{recon}) = \sum_{i=1}^{2b} \|v_i - W_d z_i\|_2^2$

- Low-dimensional representation이 noise에 강건하도록



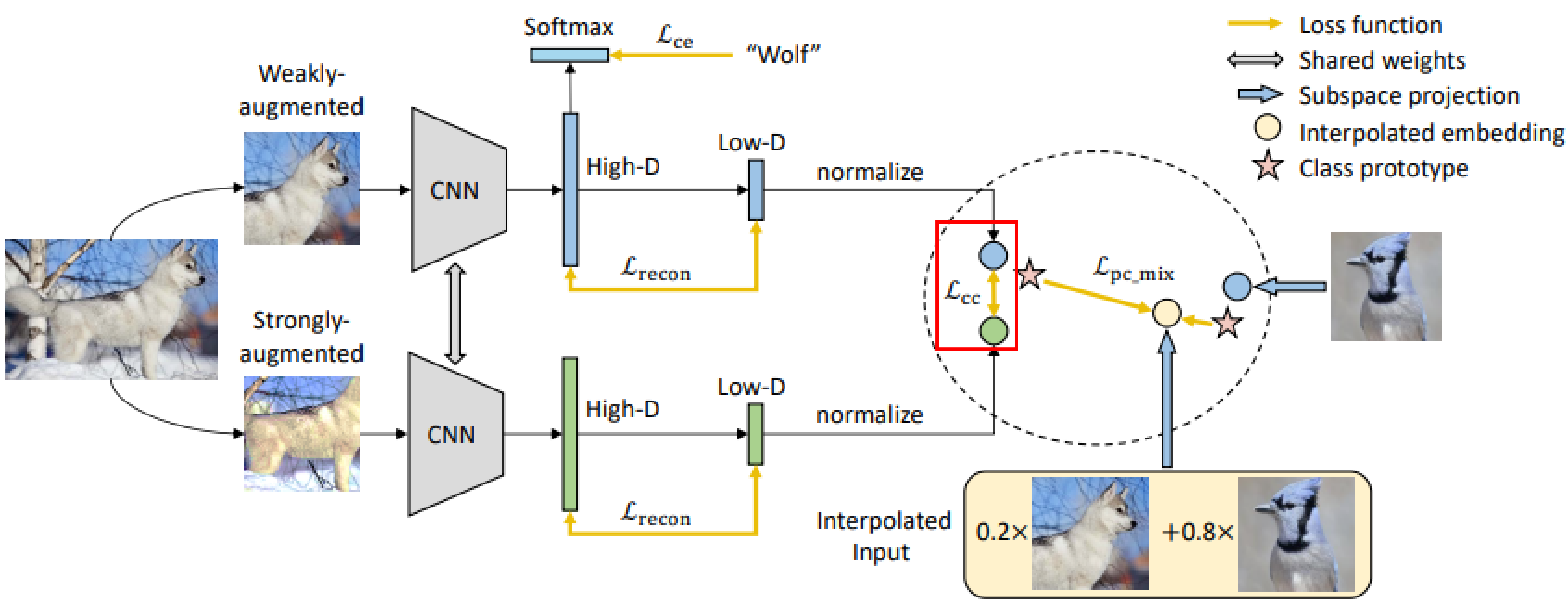
DMQA

RRL

Method

3. Consistency contrastive loss $(L_{cc}) = \sum_{i=1}^{b} -\log \frac{\exp(\hat{z}_i \cdot \hat{z}_{j(i)}/\tau)}{\sum_{k=1}^{2b} 1_{i \neq k} \exp(\hat{z}_i \cdot \hat{z}_k/\tau)}$

- 지속적인 semantic 학습 → 표현 공간에서 ID와 OOD를 지속적으로 분리



DMQA

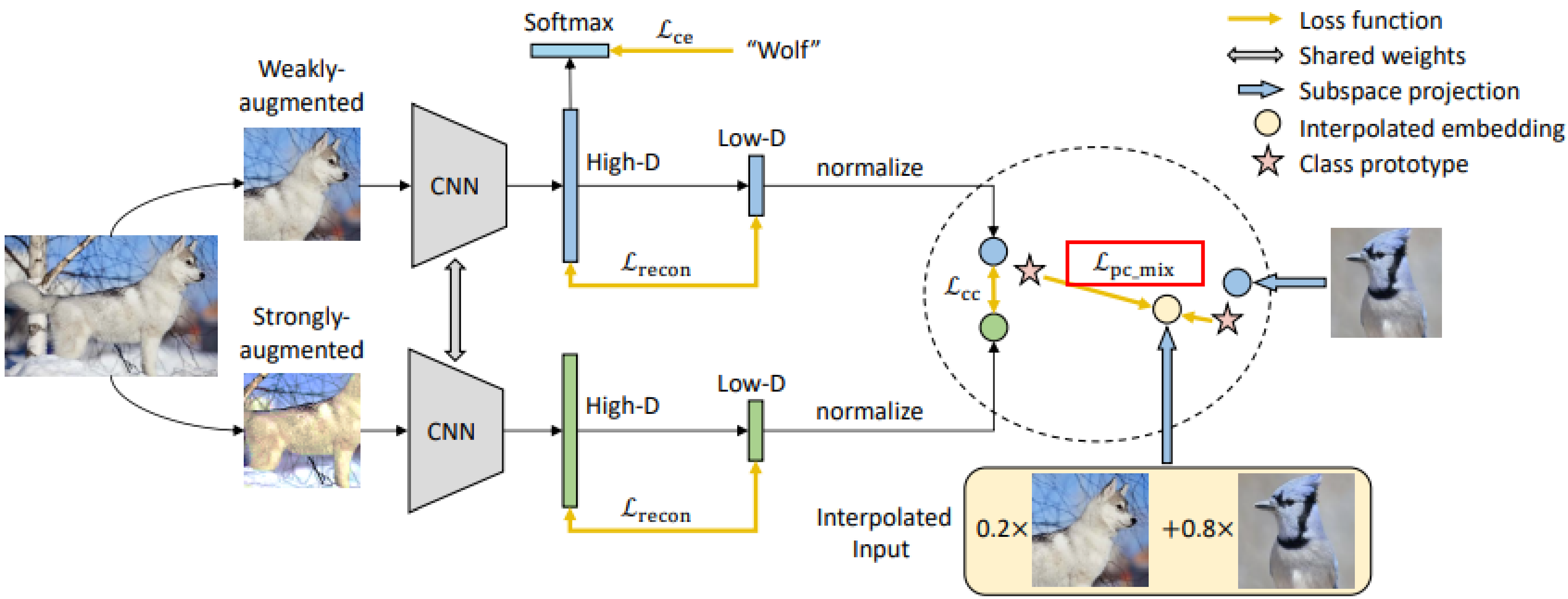
RRL

$$L_{pc} = -\log \frac{\exp(\hat{z}_i \cdot \hat{z}^{y_i}/\tau)}{\sum_{c=1}^{C} \exp(\hat{z}_i \cdot \hat{z}^{c}/\tau)}$$

Method

4. Mixup prototypical contrastive loss $(L_{pc\_mix}) = \sum_{i=1}^{b} \lambda L_{pc}(\hat{z}_i^m, y_i) + (1-\lambda) L_{pc}(\hat{z}_i^m, y_{m(i)})$

- Noise에 강건한 mixup 방식 사용
- Class 정보까지 활용하여 semantic 학습



DMQA

RRL

$$w_{ij}^t = \frac{\exp(\hat{z}_i^t \cdot \hat{z}_j^t / \tau)}{\sum_{j=1}^{k} \exp(\hat{z}_i^t \cdot \hat{z}_j^t / \tau)}$$

Method

❖ Noise cleaning with smooth neighbors

- Warm up 이후, KNN 기반의 label smoothing
- Weakly-supervised subset을 생성하여 해당 subset에서만 L_{ce}, L_{pc_mix} 계산
  - ➢ Unsupervised loss인 L_{cc}, L_{recon}은 모든 데이터에 대해서 계산

현재 epoch(t)에서 softmax prediction

$$q_i^t = \frac{1}{2} p_i^t + \frac{1}{2} \sum_{j=1}^{k} w_{ij}^t q_j^{t-1}$$

이전 epoch(t-1)에서의 soft pseudo-label

기존 label의 confidence가 높은 집합

$$D_{ws}^t = \{x_i, y_i | q_i^t(y_i) > \eta_0\} \cup \{x_i, \hat{y}_t = \arg\max_c q_i^t(c) | \forall \max_c q_i^t > \eta_1, c \in \{1, \dots, C\}\}$$

기존 label을 대신할 class의 confidence가 높은 집합

RRL

$$w_{ij}^t = \frac{\exp(\hat{z}_i^t \cdot \hat{z}_j^t/\tau)}{\sum_{j=1}^{k} \exp(\hat{z}_i^t \cdot \hat{z}_j^t/\tau)}$$

Method

❖ Noise cleaning with smooth neighbors

- Warm up 이후, KNN 기반의 label smoothing
- Weakly-supervised subset을 생성하여 해당 subset에서만 L_{ce}, L_{pc_mix} 계산
  - ➢ Unsupervised loss인 L_{cc}, L_{recon}은 모든 데이터에 대해서 계산

현재 epoch(t)에서 softmax prediction

$$q_i^t = \frac{1}{2} p_i^t + \frac{1}{2} \sum_{j=1}^{k} w_{ij}^t q_j^{t-1}$$

이전 epoch(t-1)에서의 soft pseudo-label

OOD로 의심되는 sample들은 학습에 사용하지 말자

기존 label의 confidence가 높은 집합

$$D_{ws}^t = \{x_i, y_i | q_i^t(y_i) > \eta_0\} \cup \{x_i, \hat{y}_t = \arg\max_c q_i^t(c) | \forall \max_c q_i^t > \eta_1, c \in \{1, \dots, C\}\}$$

기존 label을 대신할 class의 confidence가 높은 집합

RRL

Method

❖ Noise cleaning with smooth neighbors

- Pseudo-label의 정확도가 점진적으로 향상
- Subset의 규모가 커지더라도, 내부 noise 비율은 작게 유지 → 점점 noise에 덜 오염됨

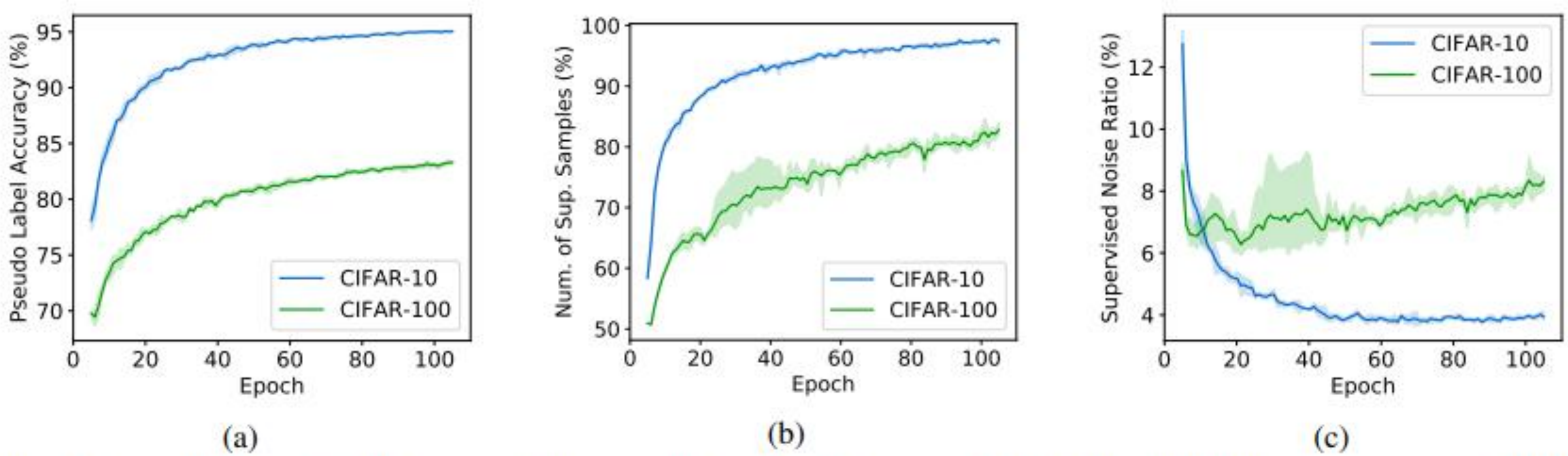


Figure 3. Curriculum learned by the proposed label correction method for training on CIFAR datasets with 50% sym. noise. (a) Accuracy of pseudo-labels *w.r.t* to clean training labels. Our method effectively cleans a majority of the label noise. (b) Number of samples in the weakly-supervised subset $\mathcal{D}^t_{ws}$. As the pseudo-labels become more accurate, more samples are used to compute the supervised losses. (c) Label noise ratio in the weakly-supervised subset, which maintains at a low level even as the size of the subset grows.

RRL

Method

❖ **Algorithm**

Warm up (lines 2–13)

Train with D^t_{ws} (lines 15–19)

Algorithm 1: Pseudo-code for our method.

```
1  Input: noisy training data D = {(x_i, y_i)}_{i=1}^n, model
     parameters θ.
2  for t ← 0 to t_0 − 1 do            // learn from noisy
     labels for t_0 epochs (warm-up)
3      {ẑ_i}_{i=1}^n = {f_θ(x_i)}_{i=1}^n
           // get normalized low-dimentional
       embeddings for all images
4      {ẑ^c}_{c=1}^C = Calculate-Prototype({ẑ_i, y_i}_{i=1}^n)
           // calculate class prototypes
5      for {(x_i, y_i)}_{i=1}^{2b} in D do          // load a
       minibatch
6          ẑ_i = f_θ(x_i)          // obtain normalized
           low-dimensional embeddings
7          λ ~ Beta(α, α)          // sample a mixup
           weight from a beta distribution
8          x_i^m = λx_i + (1 − λ)x_{m(i)}         // generate
           virual training samples
9          ẑ_i^m = f_θ(x_i^m) // obtain embeddings for
           virtual samples
10         L = Σ_{i=1}^b L_ce(x_i, y_i) + Σ_{i=1}^{2b} (ω_cc L_cc(ẑ_i) +
           ω_pc L_pc_mix(ẑ_i^m, y_i, λ) + ω_recon L_recon(x_i, ẑ_i))
11         θ = SGD(L, θ)          // compute loss and
           update model parameters
12     end
13 end
14 for t ← t_0 to MaxEpoch do            // learn from
     psuedo-labels
15     {ẑ_i^t, p_i^t}_{i=1}^n = {f_θ(x_i)}_{i=1}^n
           // get embeddings and softmax
       predictions for all images
16     q_i^t = ½ p_i^t + ½ Σ_{j=1}^k w_ij^t q_j^{t−1} ,  q_i^{t_0−1} = p_i^{t_0}
           // aggregate information from top-k
       neighbors to generate soft labels
17     D_ws^t = {x_i, y_i | q_i^t(y_i) > η_0} ∪ {x_i, ŷ_i^t =
       arg max_c q_i^t(c) | ∀ max_c q_i^t(c) > η_1, c ∈ {1, ..., C}}
           // construct a subset containing clean
       samples and pseudo-labeled samples
18     Repeat line 4-12, but only use samples from D_ws^t
         to compute ẑ^c, L_ce, L_pc_mix.
19 end
```

RRL

Experiments

❖ Main results

- Symmetric / asymmetric noise 상황에서 가장 우수한 성능 도출

Dataset Noise type	CIFAR-10 Sym 20%	 Sym 50%	 Asym 40%	CIFAR-100 Sym 20%	 Sym 50%
Cross-Entropy [18]	82.7	57.9	72.3	61.8	37.3
Forward [30]	83.1	59.4	83.1	61.4	37.3
Co-teaching+ [42]	88.2	84.1	-	64.1	45.3
Mixup [44]	92.3	77.6	-	66.0	46.6
P-correction [41]	92.0	88.7	88.1	68.1	56.4
MLNT [19]	92.0	88.8	88.6	67.7	58.0
M-correction [1]	93.8	91.9	86.3	73.4	65.4
DivideMix [18]	95.0	93.7	91.4	74.8	72.1
ELR [23] (reproduced)	94.7±0.1	93.5±0.2	91.7±0.9	75.3±0.2	71.3±0.3
DivideMix (reproduced)	95.1±0.1	93.6±0.2	91.3±0.8	75.1±0.2	72.1±0.3
Ours (classifier)	95.8±0.1	94.3±0.2	91.9±0.8	79.1±0.1	74.8±0.4
Ours (knn)	**95.9**±0.1	**94.5**±0.1	**92.4**±0.9	**79.4**±0.1	**75.0**±0.4

DMQA

RRL

Experiments

❖ Main results

- OOD까지 추가된 문제 상황에서도 가장 우수한 성능 도출

Figure 4. Examples of input noise injected to CIFAR-10. using the learned embeddings demonstrates high robustness to input noise.

Input noise	CE	Iterative [38]	GCE [45]	DivideMix [18]	Ours (cls.)	Ours (knn)
+ CIFAR-100 20k	53.6	87.2	87.3	89.0	91.5	**93.1**±0.3
+ SVHN 20k	58.1	88.6	88.8	91.9	93.3	**93.9**±0.2
Image corruption	53.8	87.7	87.9	89.8	91.4	**91.6**±0.2

Table 2. Comparison with state-of-the-art methods on CIFAR-10 dataset with label noise (50% symmetric) and input noise (OOD images or corrupted images). Numbers indicate average test accuracy (%) over last 10 epochs. We report results over 3 independent runs with randomly-generated noise. We re-run previous methods using publicly available code with the same data and model as ours.

RRL

Experiments

❖ 학습에 따른 표현 공간 시각화

- 학습이 지속될 수록 embedding space에서 OOD를 잘 분리

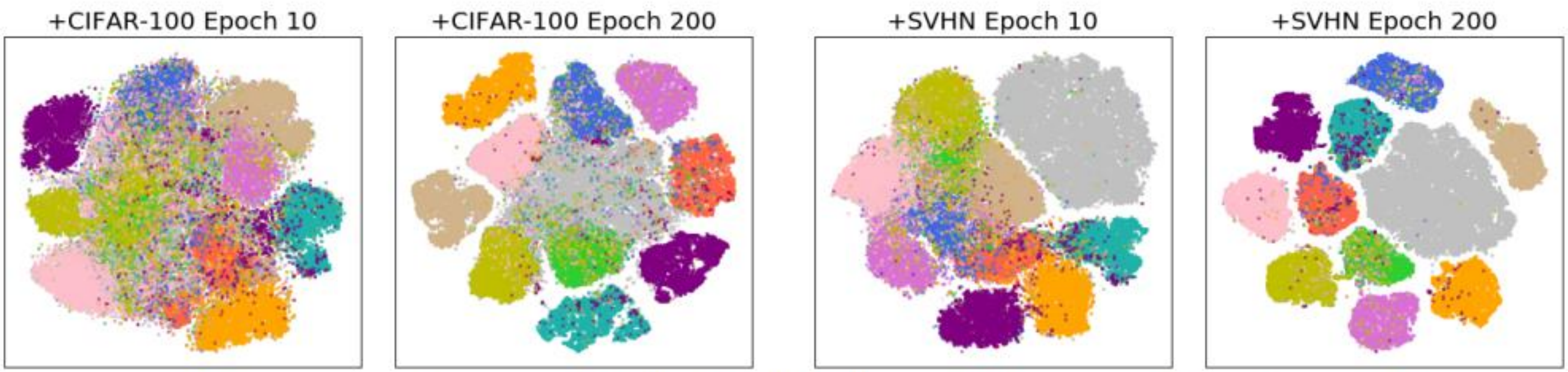


Figure 5. t-SNE visualization of low-dimensional embeddings for CIFAR-10 images (color represents the true class) + OOD images (gray points) from CIFAR-100 or SVHN. The model is trained on noisy CIFAR-10 ($50k$ images with 50% label noise) and $20k$ OOD images with random labels. Our method can effectively learn to (1) cluster CIFAR-10 images according to their true class, despite their noisy labels; (2) separate OOD samples from in-distribution samples, such that their harm is reduced.

DMQA

RRL

Experiments

❖ Real-world noisy datasets results

- 실제 noisy 벤치마크에서도 가장 우수한 성능을 보임
- Noise cleaning 방식의 효과도 같이 확인

Test dataset	WebVision		ILSVRC12	
Accuracy (%)	top1	top5	top1	top5
Forward [30]	61.1	82.7	57.4	82.4
Decoupling [27]	62.5	84.7	58.3	82.3
D2L [25]	62.7	84.0	57.8	81.4
MentorNet [14]	63.0	81.4	57.8	79.9
Co-teaching [6]	63.6	85.2	61.5	84.7
INCV [3]	65.2	85.3	61.0	85.0
ELR [23]	76.3	91.3	68.7	87.8
DivideMix [18]	75.9	90.1	73.3	89.2
Ours (w/o noise cleaning)	75.5	90.2	72.0	90.0
Ours (classifier)	76.3	**91.5**	73.3	**91.2**
Ours (knn)	**77.8**	91.3	**74.4**	90.9

Method	CE	Forward	Joint-Opt	ELR	MLNT	MentorMix	SL	DivideMix	Ours (cls.)	Ours (knn)
Accuracy	69.21	69.84	72.16	72.87	73.47	74.30	74.45	74.48	74.84	**74.97**

Table 6. Comparison with state-of-the-art methods on Clothing1M dataset. Results for previous methods are directly copied from corresponding papers. We report results for ELR and DivideMix without model ensemble.

RRL

Experiments

❖ Ablation study

- **Prototype contrastive loss**가 가장 중요한 요소임을 확인

	CIFAR-10 Sym 50%	+ CIFAR-100 20k	+ Image Corruption	CIFAR-100 Sym 50%
w/o $\mathcal{L}_{\text{pc\_mix}}$	85.9 (86.1)	79.7 (81.5)	81.6 (81.7)	65.6 (65.9)
w/o $\mathcal{L}_{\text{cc}}$	93.7 (93.8)	91.3 (91.5)	89.4 (89.5)	71.9 (71.8)
w/o $\mathcal{L}_{\text{recon}}$	93.3 (94.0)	90.7 (92.9)	90.2 (91.0)	73.2 (73.9)
w/o mixup	89.5 (89.9)	85.4 (87.0)	84.7 (84.9)	69.3 (69.7)
w/ standard aug.	94.1 (94.3)	90.8 (92.9)	90.5 (90.7)	74.5 (75.0)
DivideMix	93.6	89.0	89.8	72.1
Ours	94.3 (94.5)	91.5 (93.1)	91.4 (91.6)	74.8 (75.0)

DSOS

Paper

❖ **Addressing Out-of-Distribution Label Noise in Webly-Labelled Data (2022, WACV)**

- 기존 label correction의 한계를 지적하며 clean / ID noise / OOD를 잘 분리할 수 있는 metric 제시

Addressing out-of-distribution label noise in webly-labelled data

Paul Albert, Diego Ortego, Eric Arazo, Noel E. O'Connor, Kevin McGuinness

School of Electronic Engineering,
Insight SFI Centre for Data Analytics, Dublin City University (DCU)
paul.albert@insight-centre.org

Abstract

A recurring focus of the deep learning community is towards reducing the labeling effort. Data gathering and annotation using a search engine is a simple alternative to generating a fully human-annotated and human-gathered dataset. Although web crawling is very time efficient, some of the retrieved images are unavoidably noisy, i.e. incorrectly labeled. Designing robust algorithms for training on noisy data gathered from the web is an important research perspective that would render the building of datasets easier. In this paper we conduct a study to understand the type of label noise to expect when building a dataset using a search engine. We review the current limitations of state-of-the-art methods for dealing with noisy labels for image classification tasks in the case of web noise distribution. We propose a simple solution to bridge the gap with a fully clean dataset using Dynamic Softening of Out-of-distribution Samples (DSOS), which we design on corrupted versions of the CIFAR-100 dataset, and compare against state-of-the-art algorithms on the web noise perturbated MiniImageNet and Stanford datasets and on real label noise datasets: WebVision 1.0 and Clothing1M. Our work is fully reproducible https://git.io/JKGcj.

1. Introduction

Deep neural networks (DNNs) are now the standard approach for accurately solving image classification tasks [26, 48]. However, their principal drawback is the large amount of labeled examples required for training. There exist numerous alternatives to deal with the limited availability of labels, such as but not limited to, semi-supervised learning [1, 3, 4], self-supervised learning [8, 5] and robust training on automatically annotated datasets [23, 12]. This paper focuses on the latter.

Designing robust algorithms to train image classification DNNs in the presence of label noise is an important focus for the community [36]; these enable better adaptation of current DNN solutions to real-world problems where extensive curated datasets are unavailable or too expensive to build. Controlled label noise datasets are then often created by synthetically introducing label corruptions in the CIFAR-100 [17] comparison benchmark. Although good noise robustness is shown on these artificial datasets, web label noise has proven that these solutions generalize poorly to more realistic scenarios and can, in specific cases, be outperformed by robust data augmentation strategies such as mixup [12, 27].

We hypothesize that the main limitation for the correction of label noise in web crawled datasets comes from a common assumption made by most label noise robust algorithms [21, 30, 29, 41] where the true labels for noisy samples lie inside the label set, i.e. the label noise is *in-distribution* (ID). Conversely, we hypothesize that the label noise present in web crawled datasets is predominantly *out-of-distribution* (OOD), meaning the real labels for noisy samples cannot be inferred from the distribution. To confirm our hypothesis, we conduct a small but representative survey on the WebVision 1.0 dataset [23] to identify the type of noise one can expect in automatically annotated datasets crawled from the web. We then build and validate the DSOS method on controlled corrupted versions of the CIFAR-100 dataset [17] where ID noise is introduced using symmetric label flipping and where we use the ImageNet32 [6] dataset to introduce OOD noise. We compare with state-of-the-art label noise algorithms on multiple real-world open-source web-crawled datatsets including corrupted versions of the miniImageNet [39] and Stanford Cars [16] datasets provided by Jiang *et al.* [12], the mini-WebVision dataset [23], and the Clothing1M [44] dataset. We observe that noisy OOD samples can be leveraged to improve network generalization by enforcing dynamically softening of labels tending to a uniform distribution [20] rather than discarding them.

This paper's contributions are:

1. We conduct a representative survey over the type of noise to be expected when constructing a dataset using web queries.

1

Albert, P., Ortego, D., Arazo, E., O'Connor, N. E., & McGuinness, K. (2022). Addressing out-of-distribution label noise in webly-labelled data. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 392-401).

DMQA

DSOS

Background

❖ 실제 noisy benchmark에는 OOD samples도 다수 존재!

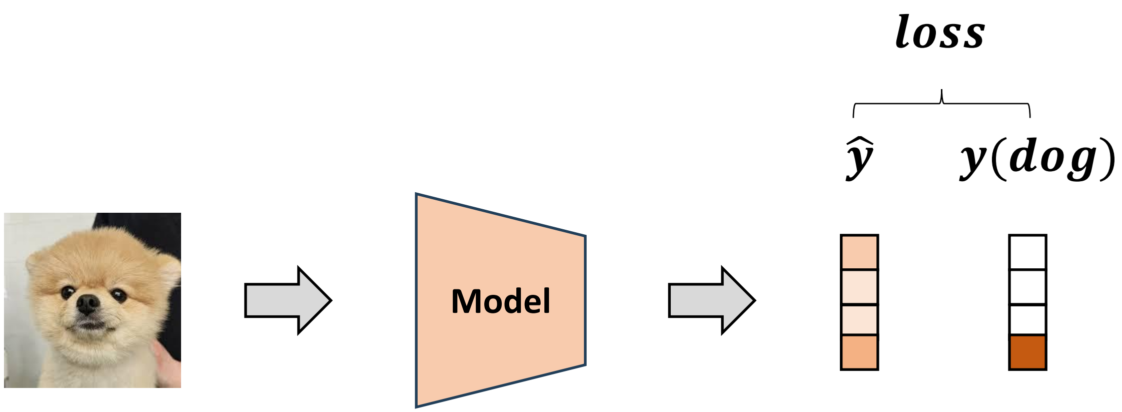
Table 1: Analysis on the noise types and ratios found in mini-WebVision. We randomly sample three subsets (S) of 2000 images and report correctly-labeled samples and in-distribution (ID) and out-of-distribution (OOD) noisy samples. Image examples are available in the supplementary material.

	S1	S2	S3	Average (%)
Correct	1441	1440	1335	1405.33 (70.30)
OOD	460	429	573	487.33 (24.38)
ID	98	130	91	106.33 (5.32)

DSOS

Method

❖ 기존 label correction 방식의 한계

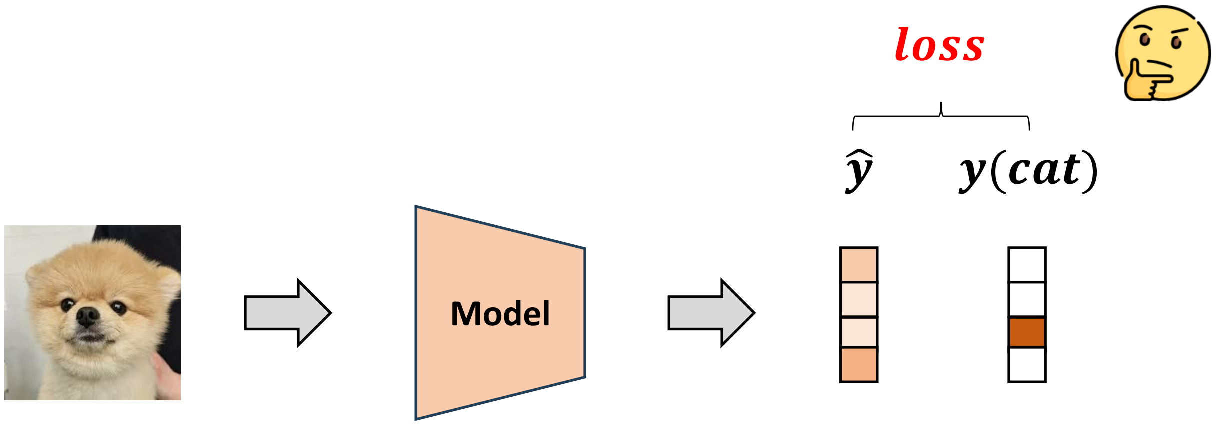


[Clean 데이터셋]

DSOS

Method

❖ 기존 label correction 방식의 한계

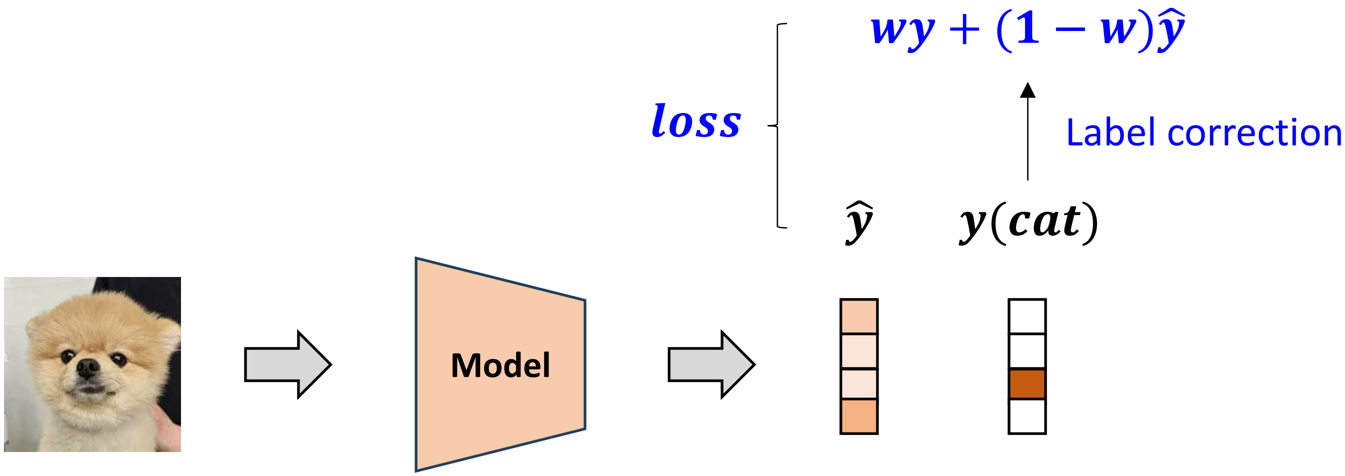


[Noisy 데이터셋]

DSOS

Method

❖ 기존 label correction 방식의 한계



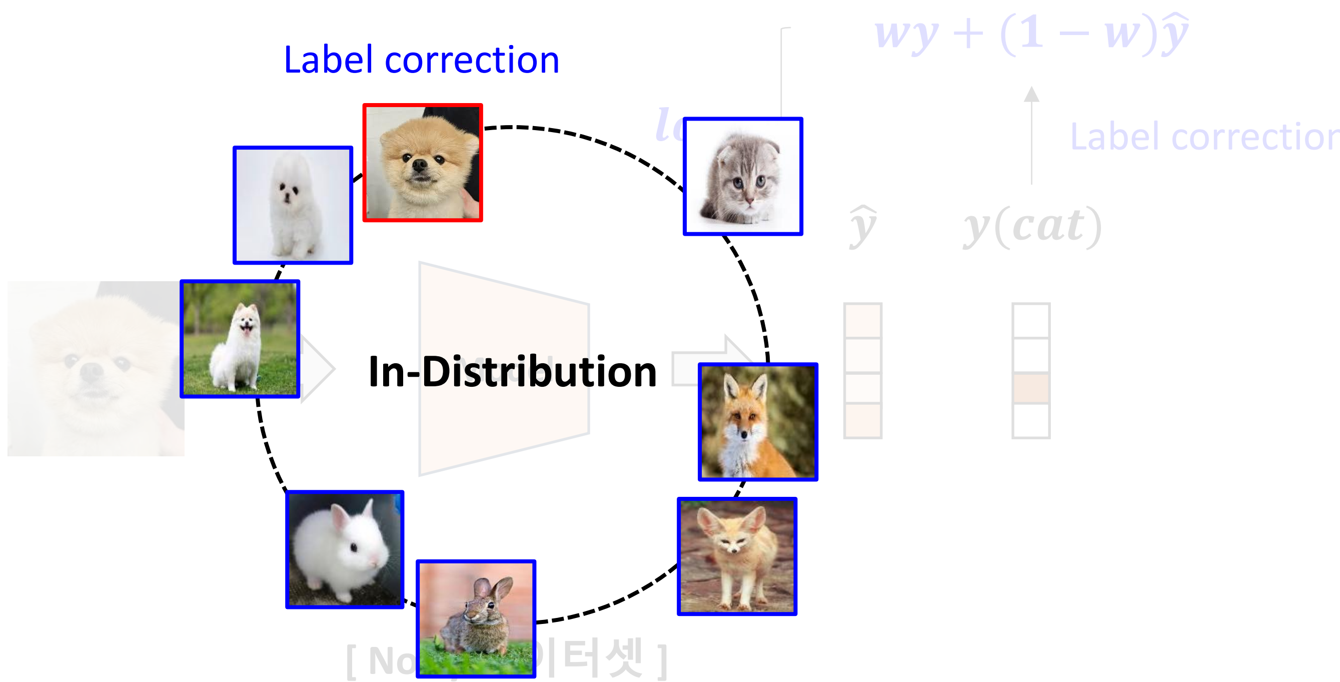
[Noisy 데이터셋]

DMQA

DSOS

Method

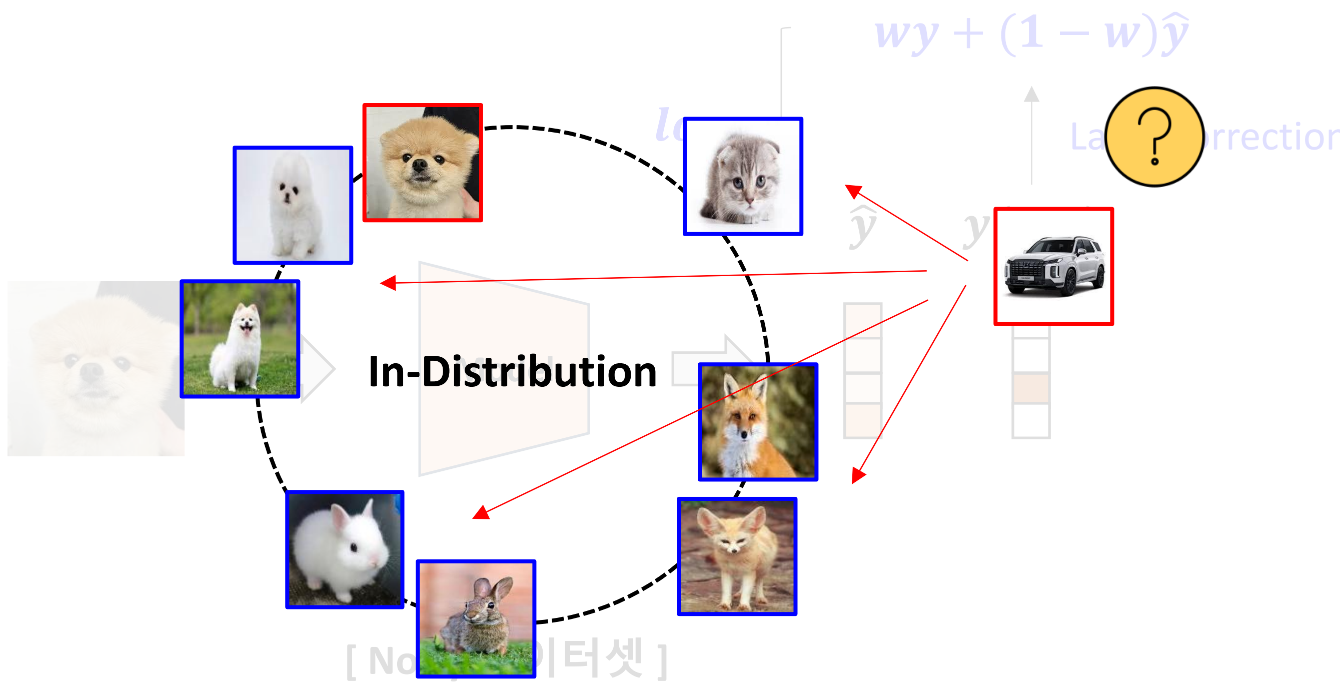
❖ 기존 label correction 방식의 한계



DSOS

Method

❖ 기존 label correction 방식의 한계



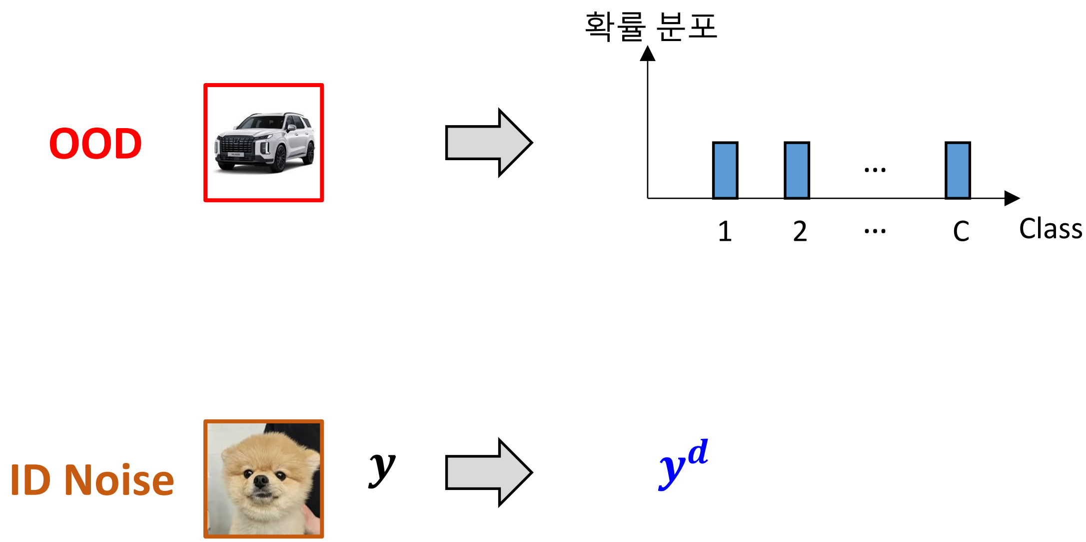
OOD는 label correction이 소용이 없다!

DMQA

DSOS

Method

❖ OOD로 판단된 sample은 uniform distribution 할당 + ID noise로 판단된 sample만 label correction



[Dynamic Softening of OOD Samples]

DMQA

DSOS

Method

❖ **Clean / ID noise / OOD 를 사전에 잘 구분하는 것이 중요**

- 새로운 **intermediate label (y_{int})**을 정의
- 실제 label과 모델 예측 값을 동시에 반영

label

prediction

$$y_{int} = \frac{y_i + \tilde{y}_i}{2}$$

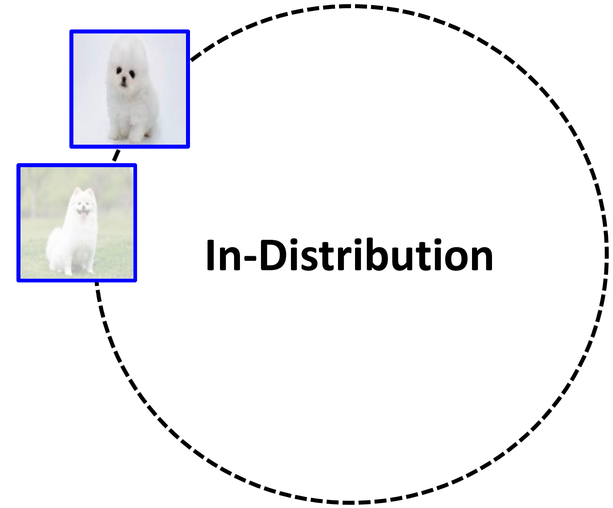
$$l_{detect} = -\log \sum_{c=1}^{C} y_{int,c}^2$$

$$y_i = [0,0,0,1]$$

$$\tilde{y}_i = [0.05,0.05,0.1,0.8]$$

$$y_{int} = [0.025,0.025,0.05,0.9]$$

$$l_{detect} = 0.206$$



DMQA

DSOS

Method

❖ **Clean / ID noise / OOD 를 사전에 잘 구분하는 것이 중요**

- 새로운 **intermediate label (y_{int})**을 정의
- 실제 label과 모델 예측 값을 동시에 반영

label

prediction

$$y_{int} = \frac{y_i + \tilde{y}_i}{2}$$

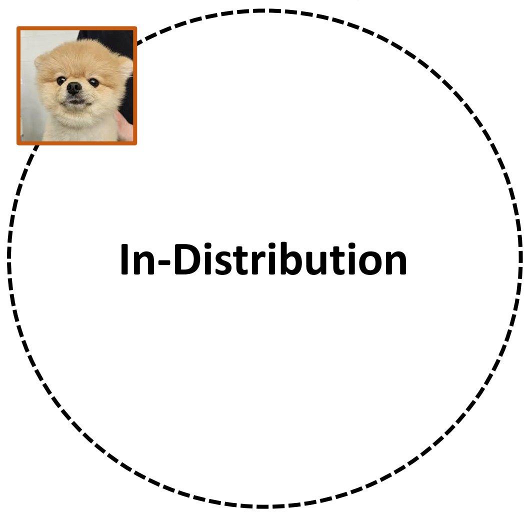
$$l_{detect} = -\log \sum_{c=1}^{C} y_{int,c}^2$$

$$y_i = [0,0,1,0]$$

$$\tilde{y}_i = [0.05,0.05,0.1,0.8]$$

$$y_{int} = [0.025,0.025,0.55,0.4]$$

$$l_{detect} = 0.768$$



DMQA

DSOS

Method

❖ **Clean / ID noise / OOD 를 사전에 잘 구분하는 것이 중요**

- 새로운 **intermediate label (y_{int})**을 정의
- 실제 label과 모델 예측 값을 동시에 반영

label

prediction

$$y_{int} = \frac{y_i + \tilde{y}_i}{2}$$

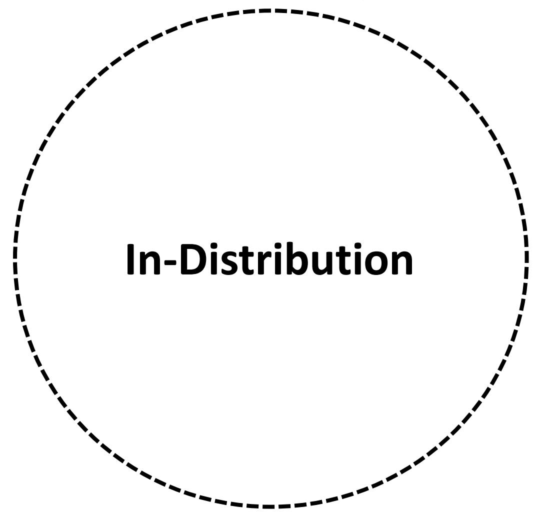
$$l_{detect} = -\log \sum_{c=1}^{C} y_{int,c}^2$$

$$y_i = [0,0,1,0]$$

$$\tilde{y}_i = [0.33,0.33,0.01,0.33]$$

$$y_{int} = [0.165,0.165,0.505,0.165]$$

$$l_{detect} = 1.089$$



DMQA

DSOS

$$l_e = -\frac{1}{N}\sum_{i=1}^{N} \tilde{v}_i \sum_{i=1}^{N} h(x_i)\log(h(x_i))$$

Method

- ID Noise sample은 예측 값으로 label correction & softening
- OOD는 uniform한 확률 분포를 가지도록 softening

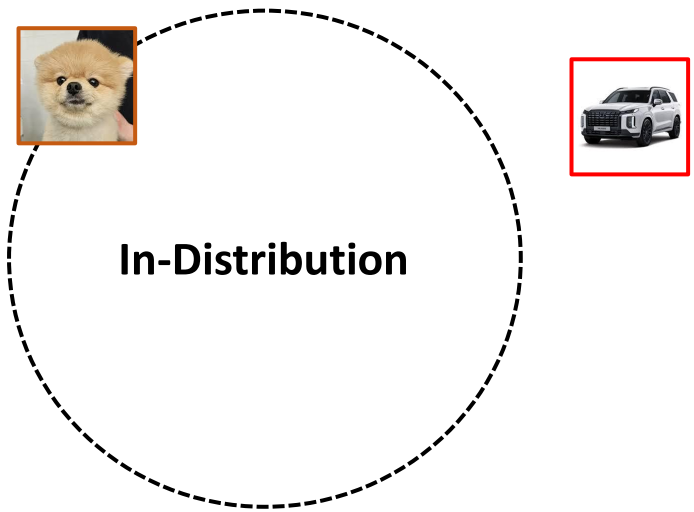
$$\begin{cases} Clean \rightarrow 0 \\ ID\ noise \rightarrow 1 \end{cases}$$

$$y_i^b = (1-\tilde{u}_i)y_i + \tilde{u}_i\tilde{y}_i^t$$

$$\begin{cases} OOD \rightarrow 0 \\ ID \rightarrow 1 \end{cases}$$

$$y_i^d = \frac{\exp\frac{\tilde{v}_i y_i^b}{\alpha}}{\sum_{c=1}^{C}\exp\frac{\tilde{v}_i y_{i,c}^b}{\alpha}}$$

$$\boldsymbol{L} = -\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}^{d^T}\log\left(\boldsymbol{h}(\boldsymbol{x}_i)\right) + \gamma \boldsymbol{l}_e$$



DSOS

Experiments

❖ Main results

- Both (ID & OOD correction)를 적용할 때, 가장 우수한 성능을 보임

Table 3: DSOS for mitigating ID and OOD noise on CIFAR-100 corrupted with ImageNet32 images. We run each algorithm with the exact same noise corruption. We report best and last accuracy (best/last).

ρ	ψ	CE	M	DB	ELR	EDM	JoSRC	DSOS		
								ID	OOD	both
0.2	0.2	63.68/55.52	66.71/62.52	65.61/65.61	63.90/63.72	65.11/64.49	67.37/64.17	68.09/67.78	69.37/69.37	70.54/70.54
0.4	0.2	58.94/44.31	59.54/53.16	54.79/54.42	57.16/56.91	55.65/54.49	61.70/61.37	60.12/59.32	62.34/61.03	62.49/62.05
0.6	0.2	46.02/26.03	42.87/40.39	42.50/42.50	31.20/29.55	28.51/10.47	37.95/37.11	46.10/42.93	46.54/40.23	49.98/49.14
0.4	0.4	41.39/18.45	38.37/33.85	35.90/35.90	22.85/21.63	24.15/01.62	41.53/41.44	40.94/35.89	42.53/39.76	43.69/42.88

DMQA

DSOS

Experiments

❖ Ablation study

- ID noise: 0.2 / OOD: 0.2
- **OOD softening까지 적용하는 것**이 가장 impact 있음을 확인

	Best	Last
CE	63.68	55.52
+ mixup	66.71	62.52
+ Entropy regularization	67.27	63.04
+ Batch normalization tuning	67.56	65.69
+ In-distribution bootstrapping	68.09	67.78
+ Out-of-distribution softening	70.54	70.54

DSOS

Experiments

❖ Real-world noisy datasets results

Table 5: Comparison of DSOS with state-of-the-art algorithms on MiniImageNet and Stanford Cars corrupted with web label noise gathered by [12] (red noise). We bold best and underline last accuracy for the best performing algorithm.

Dataset	Noise level	CE	D	SM	B	M	MN	MM	DSOS
MiniImageNet	0	70.9/68.5	71.8/65.7	71.4/68.4	71.8/68.4	72.8/72.3	71.2/68.9	74.3/73.7	**74.52**/74.10
	30	66.1/56.5	66.6/55.0	65.2/56.3	66.6/56.7	66.8/61.8	66.2/64.0	68.3/67.2	**69.84**/67.86
	50	60.9/51.7	62.1/50.01	61.3/51.3	62.6/52..5	63.2/58.4	61.7/58.0	63.3/61.8	**66.14**/65.18
	80	48.8/39.8	49.5/37.6	49.0/40.6	50.1/40.1	50.7/45.5	49.3/43.4	50.2/48.4	**55.26**/52.24
Stanford Cars	0	90.8/90.8	**92.2**/92.2	90.1/90.1	90.3/90.0	91.9/91.9	90.2/90.1	91.8/91.6	91.38/91.27
	30	80.4/80.2	87.6/87.6	82.2/81.9	83.4/83.0	85.6/85.2	81.1/80.9	87.8/87.7	**88.36**/88.14
	50	70.6/70.3	79.3/79.2	70.1/70.1	73.6/73.5	79.1/78.9	72.0/72.0	80.4/79.8	**82.04**/81.72
	80	43.3/43.0	61.8/61.8	46.4/46.4	47.4/46.7	55.7/55.4	51.0/50.9	58.6/58.6	**62.36**/62.36

Table 6: Classification accuracy for DSOS and state-of-the-art methods against methods using a unique network vs an ensemble. We train the network on the mini-Webvision dataset and test on the Imagenet 1k test set (ILSVRC12). All results except our own (DSOS) are from [24]. We bold the best results.

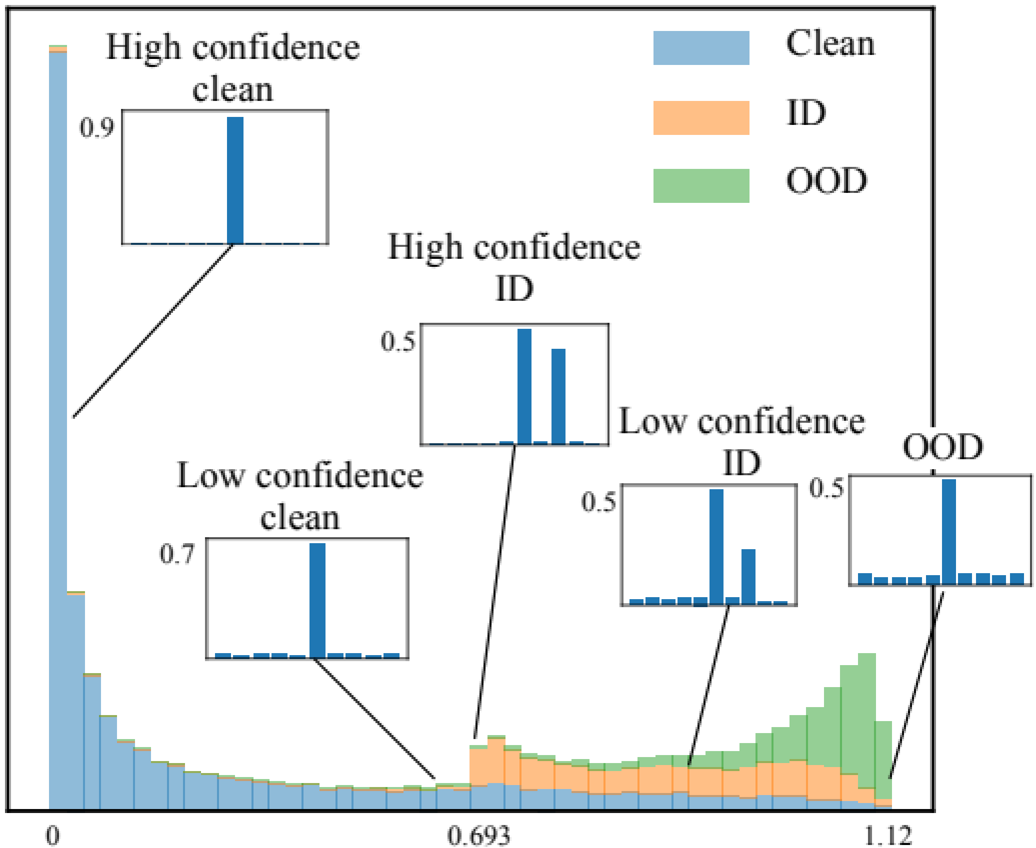
		Unique network						Ensemble of two networks		
		F	Co-T	M	MM	ELR	DSOS	DM	ELR+	DSOS
mini-WebVision	top-1	61.12	63.58	75.44	76.0	76.26	**77.76**	77.32	77.78	**78.76**
	top-5	82.68	85.20	90.12	90.2	91.26	**92.04**	91.64	91.68	**92.32**
ILSVRC12	top-1	57.36	61.48	71.44	72.9	68.71	**74.36**	75.20	70.29	**75.88**
	top-5	82.36	84.70	89.40	**91.10**	87.84	90.80	90.84	89.76	**92.36**

DMQA

DSOS

Experiments

❖ **Interpolated label (l_{detect}) 분포 시각화**



Conclusion

DMQA

Conclusion

❖ **Background**

- 일반적인 label noise learning 분야는 noise가 같은 데이터셋 분포 내에 존재한다고 가정
- 실제로 web에서 데이터를 수집할 때, **수집하고자 하는 데이터와 분포가 다른 OOD**가 존재할 수 있음

❖ **RRL**

- OOD까지 고려한 label noise learning 문제 상황을 처음으로 제시
- **Consistency & Mixup prototypical contrastive learning**을 통해 표현 공간에서 ID와 OOD를 잘 분리하며 noise에 강건하도록 학습

❖ **DSOS**

- Clean / ID noise / OOD를 효과적으로 분리하기 위한 **interpolate label**을 제시
- **Dynamic softening**을 통해 label을 보정하여 학습

References

References

❖ Li, J., Xiong, C., & Hoi, S. C. (2021). Learning from noisy data with robust representation learning. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9485-9494).

❖ Albert, P., Ortego, D., Arazo, E., O'Connor, N. E., & McGuinness, K. (2022). Addressing out-of-distribution label noise in webly-labelled data. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 392-401).

DMQA

고맙습니다

DMQA