DMQA Open Seminar

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

2025.08.01

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
- Large Language Models

❖ **Contact**
- jungmin9195@korea.ac.kr

DMQA

Contents

DMQA

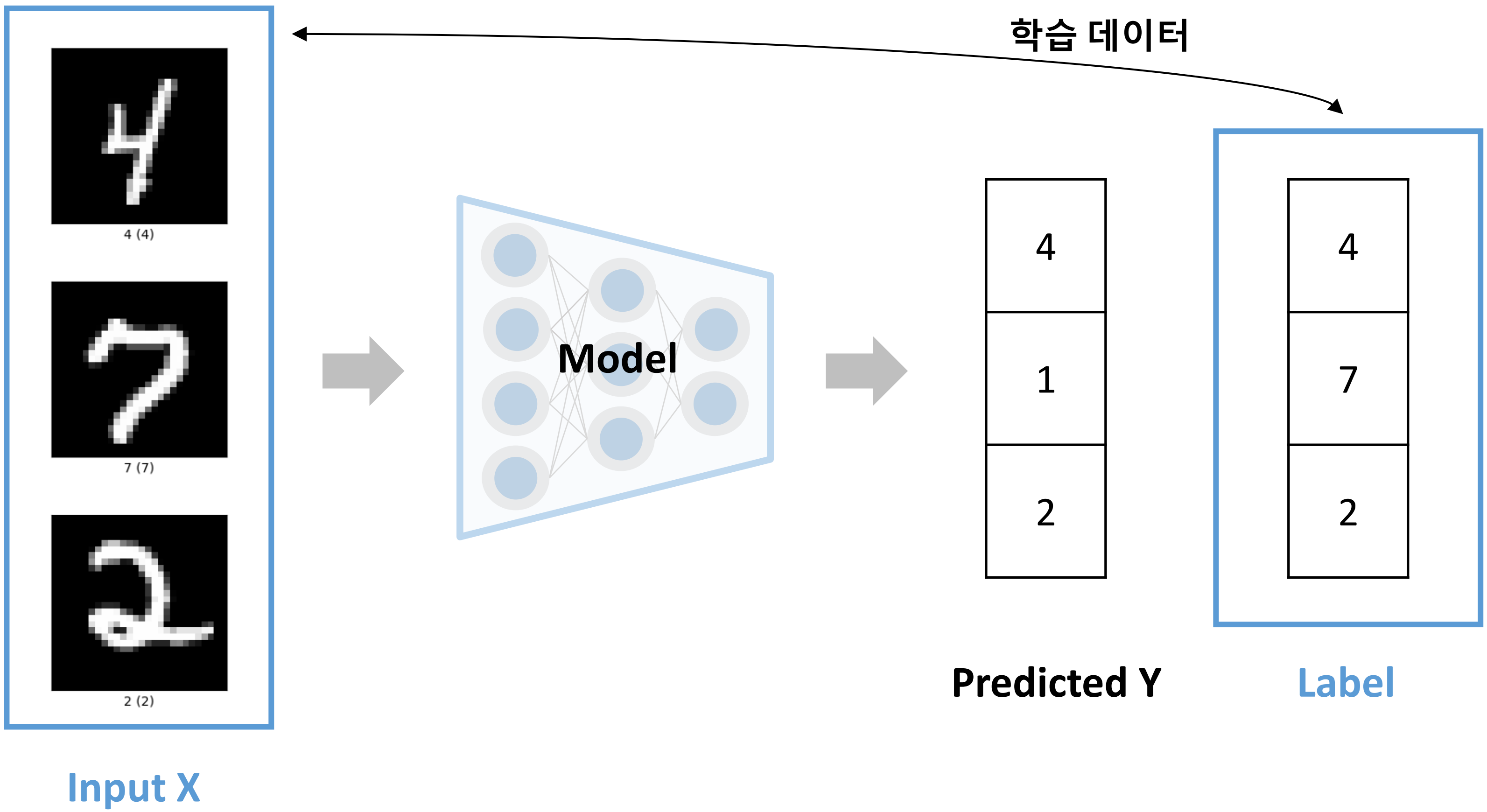
Introduction

Introduction

Noisy Label

❖ Supervised Learning

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

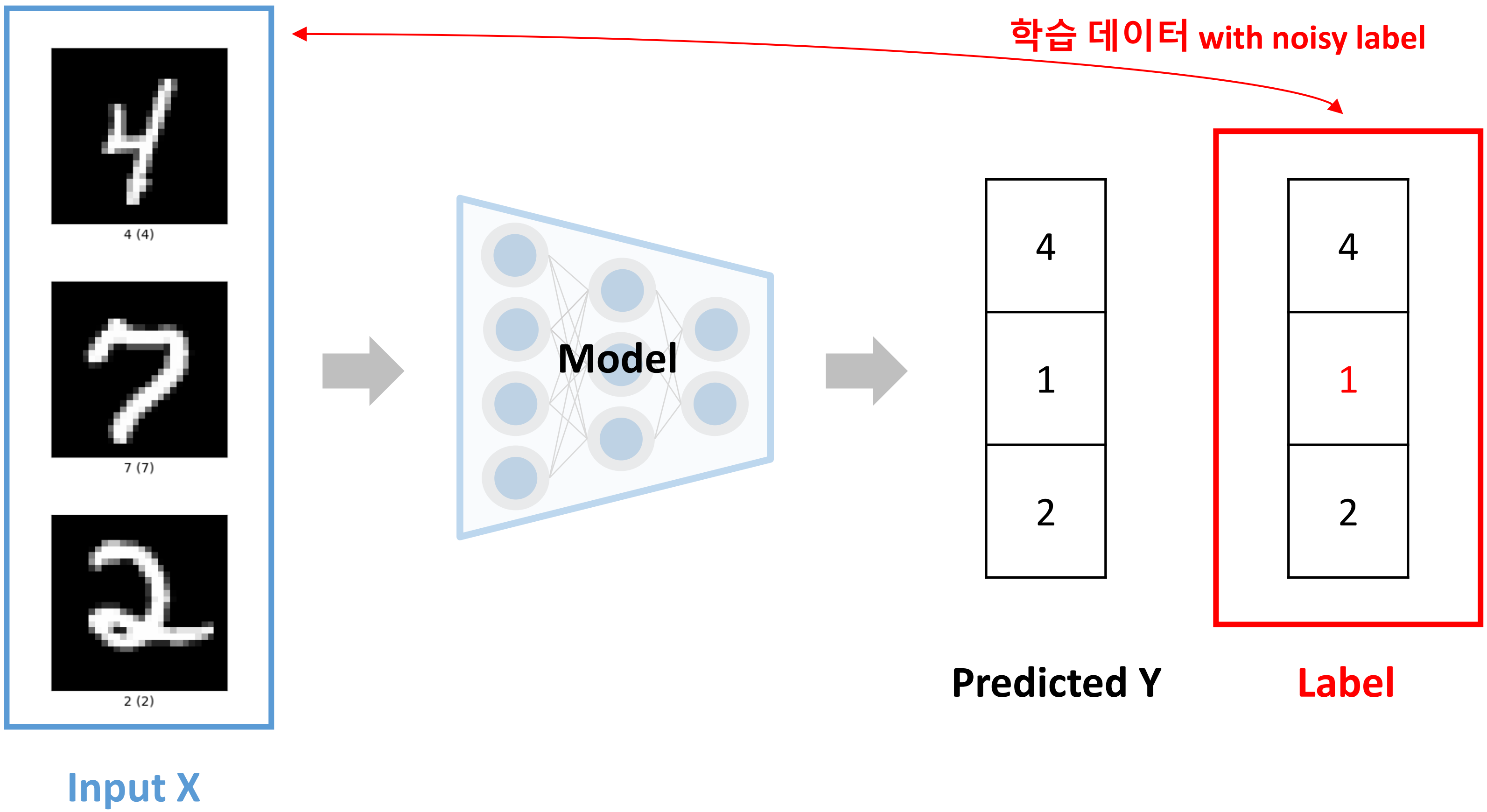


Introduction

Noisy Label

❖ Noisy Labels

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



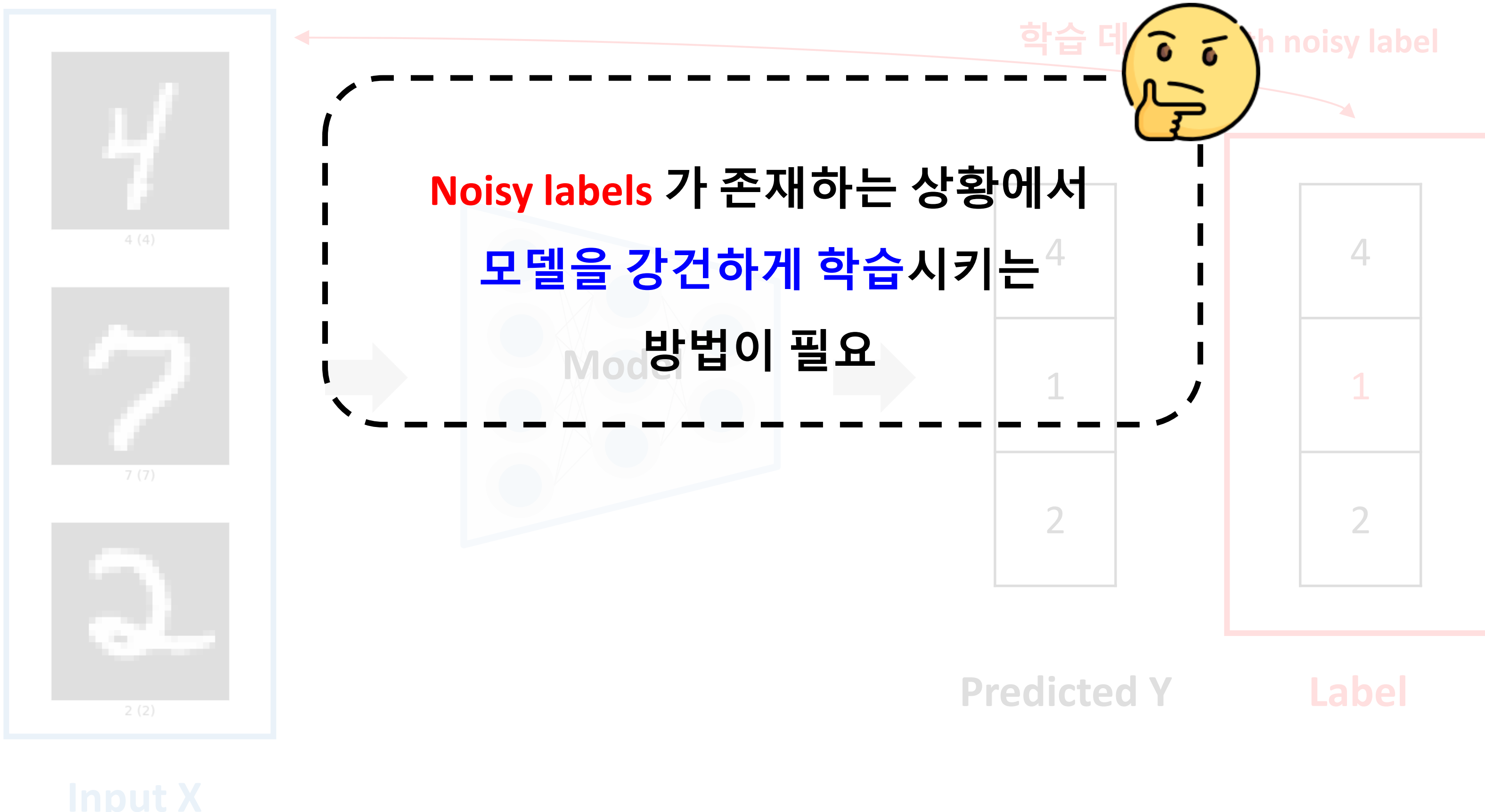
DMQA

Introduction

Noisy Label

Noisy Labels

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



Noisy labels 가 존재하는 상황에서

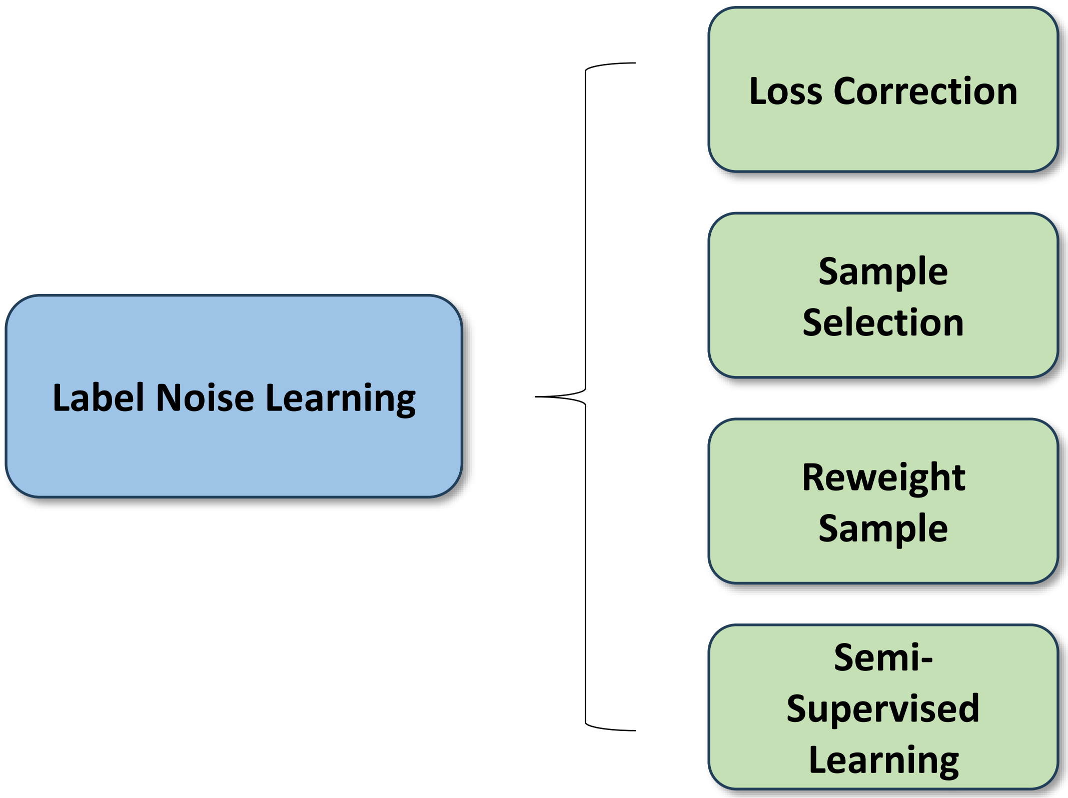
모델을 강건하게 학습시키는

방법이 필요

Introduction

Noisy Label

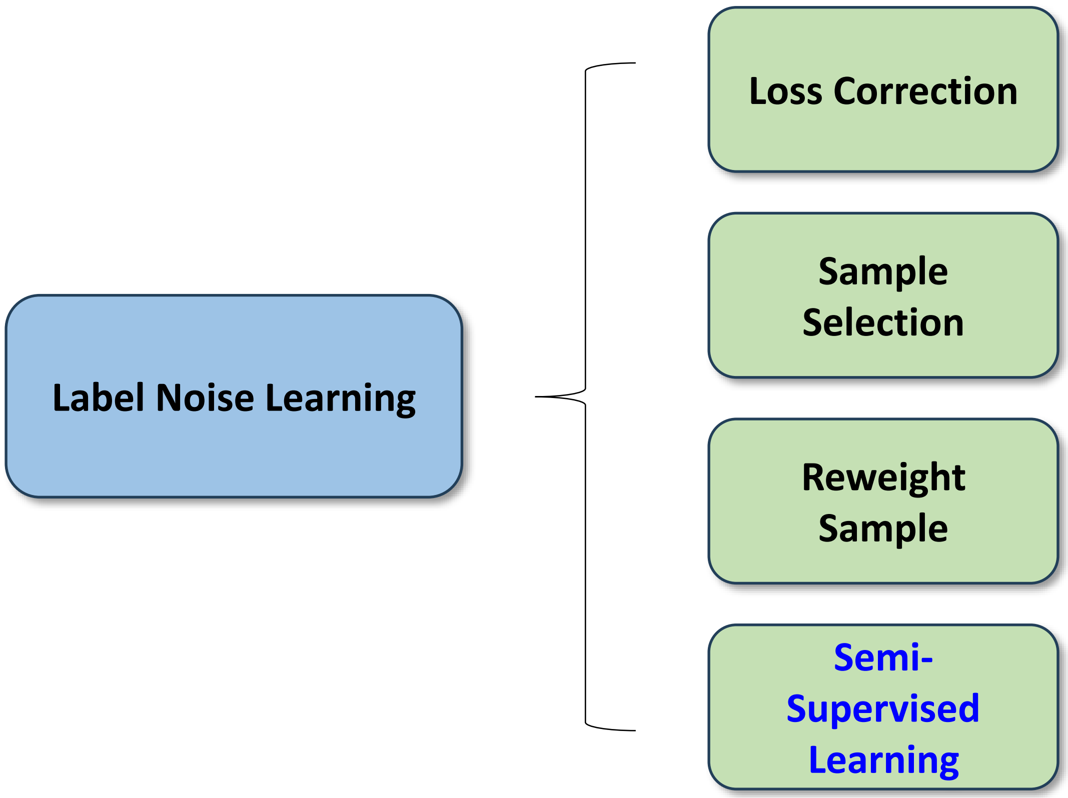
❖ Approaches for Label Noise Learning



Introduction

Noisy Label

❖ Approaches for Label Noise Learning

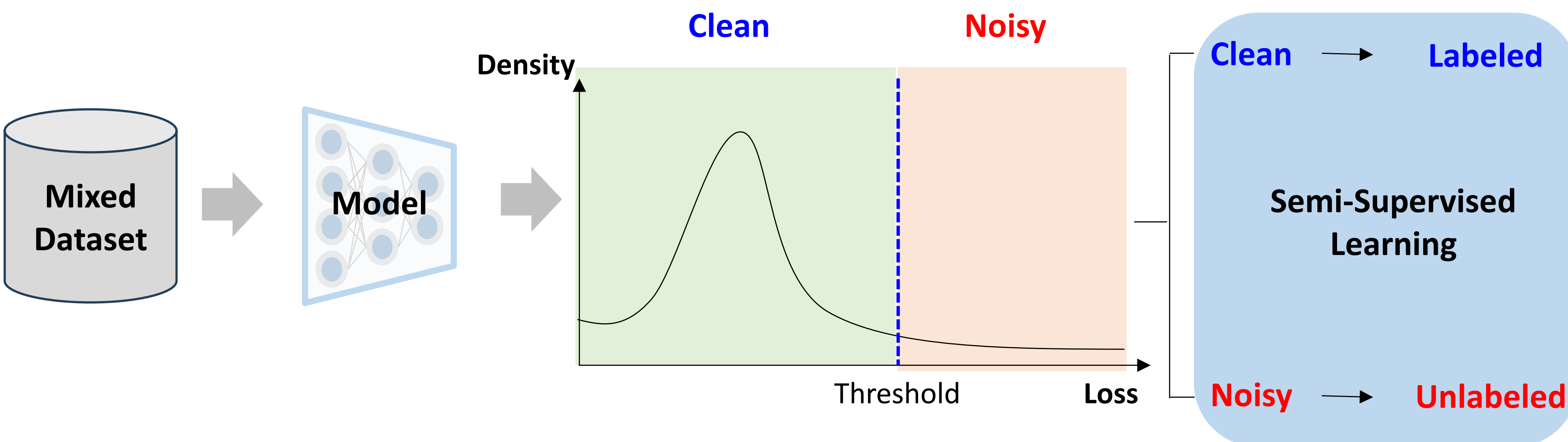


Introduction

Noisy Label

❖ **Label Noise Learning with Semi-Supervised Learning Approach**

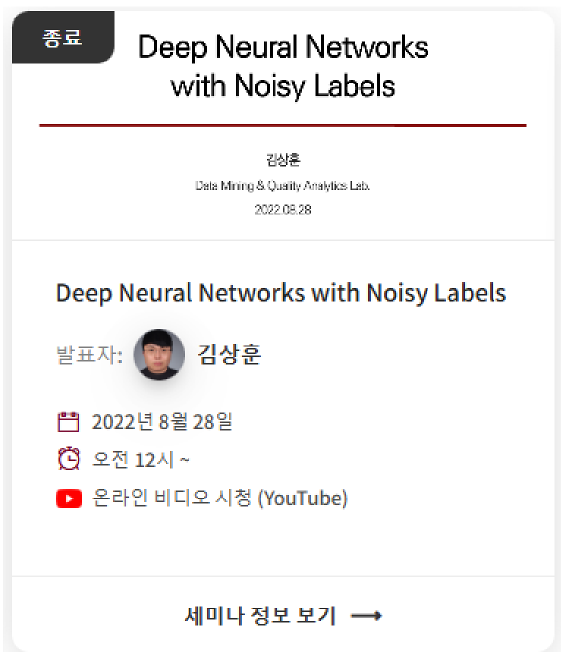
- 모델의 loss를 기준으로 특정 주기마다 clean / noisy 데이터 구분
- **Clean 데이터 → labeled 데이터** / **Noisy 데이터 → unlabeled 데이터** 취급

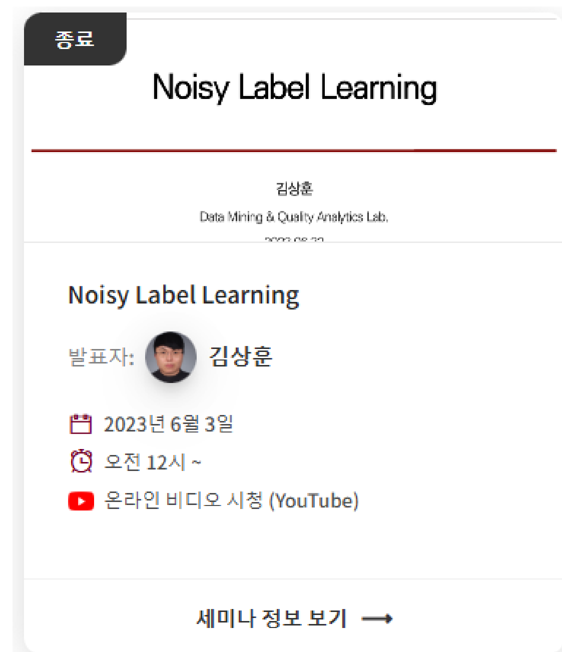


Introduction

Related Seminar

❖ Label Noise Learning





Introduction

MixUp

❖ MixMatch: A Holistic Approach to Semi-Supervised Learning (2019, Neurips)

MixMatch: A Holistic Approach to Semi-Supervised Learning

David Berthelot
Google Research
dberth@google.com

Nicholas Carlini
Google Research
ncarlini@google.com

Ian Goodfellow
Work done at Google
ian-academic@mailfence.com

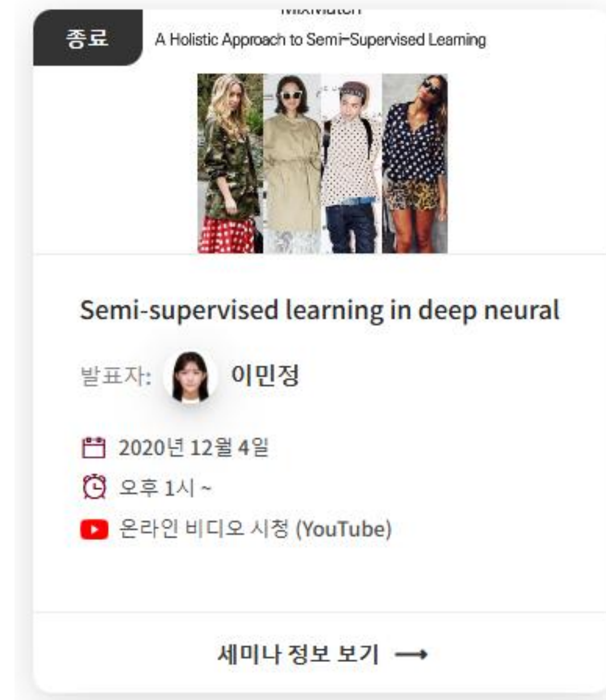
Avital Oliver
Google Research
avitalo@google.com

Nicolas Papernot
Google Research
papernot@google.com

Colin Raffel
Google Research
craffel@google.com

Abstract

Semi-supervised learning has proven to be a powerful paradigm for leveraging unlabeled data to mitigate the reliance on large labeled datasets. In this work, we unify the current dominant approaches for semi-supervised learning to produce a new algorithm, MixMatch, that guesses low-entropy labels for data-augmented unlabeled examples and mixes labeled and unlabeled data using MixUp. MixMatch obtains state-of-the-art results by a large margin across many datasets and labeled data amounts. For example, on CIFAR-10 with 250 labels, we reduce error rate by a factor of 4 (from 38% to 11%) and by a factor of 2 on STL-10. We also demonstrate how MixMatch can help achieve a dramatically better accuracy-privacy trade-off for differential privacy. Finally, we perform an ablation study to tease apart which components of MixMatch are most important for its success. We release all code used in our experiments.[1]



Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., & Raffel, C. A. (2019). Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, *32*.

DMQA

Introduction

MixUp

❖ MixUp in Semi-Supervised Learning

$\boldsymbol{\lambda'} \times$

x_1

$(\boldsymbol{1} - \boldsymbol{\lambda'}) \times$

x_2

$$\lambda \sim Beta(\alpha, \alpha)$$
$$\lambda' = \max(\lambda, 1 - \lambda)$$
$$x' = \lambda' x_1 + (1 - \lambda') x_2$$
$$p' = \lambda' p_1 + (1 - \lambda') p_2$$

DMQA

Label Noise Learning beyond DivideMix

DivideMix

Paper

❖ DivideMix: Learning with Noisy Labels as Semi-Supervised Learning (2020, ICLR)

- 준지도학습 방법론 중 MixMatch를 도입

DIVIDEMIX: LEARNING WITH NOISY LABELS AS SEMI-SUPERVISED LEARNING

Junnan Li, Richard Socher, Steven C.H. Hoi
Salesforce Research
{junnan.li,rsocher,shoi}@salesforce.com

ABSTRACT

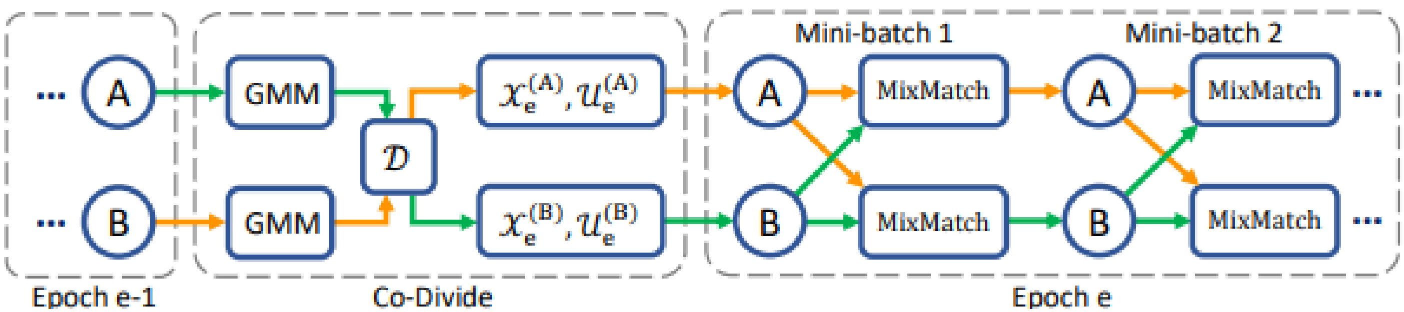
Deep neural networks are known to be annotation-hungry. Numerous efforts have been devoted to reducing the annotation cost when learning with deep networks. Two prominent directions include learning with noisy labels and semi-supervised learning by exploiting unlabeled data. In this work, we propose DivideMix, a novel framework for learning with noisy labels by leveraging semi-supervised learning techniques. In particular, DivideMix models the per-sample loss distribution with a mixture model to dynamically divide the training data into a labeled set with clean samples and an unlabeled set with noisy samples, and trains the model on both the labeled and unlabeled data in a semi-supervised manner. To avoid confirmation bias, we simultaneously train two diverged networks where each network uses the dataset division from the other network. During the semi-supervised training phase, we improve the MixMatch strategy by performing label co-refinement and label co-guessing on labeled and unlabeled samples, respectively. Experiments on multiple benchmark datasets demonstrate substantial improvements over state-of-the-art methods. Code is available at https://github.com/LiJunnan1992/DivideMix.

Li, J., Socher, R., & Hoi, S. C. (2020). Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*.

DivideMix

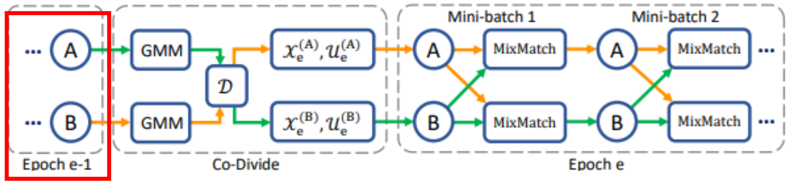
Method

❖ Overall Process



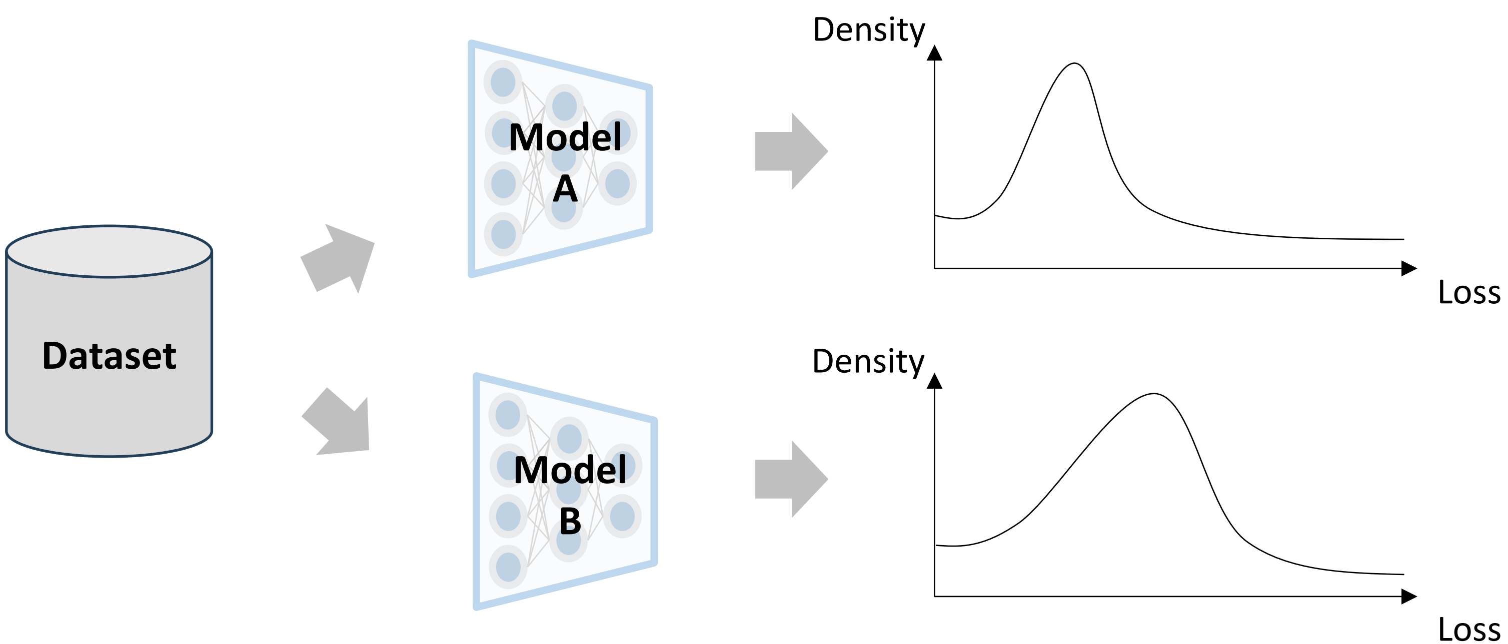
DivideMix

Method



❖ Loss Modeling

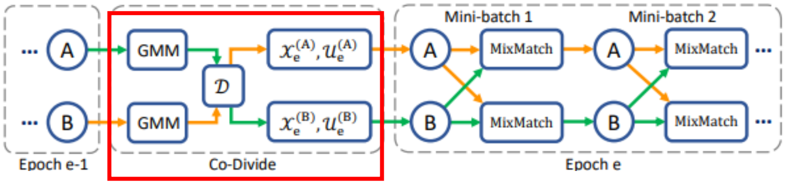
- 두 개의 네트워크를 동시에 학습하여 매 epoch마다 loss 산출



DMQA

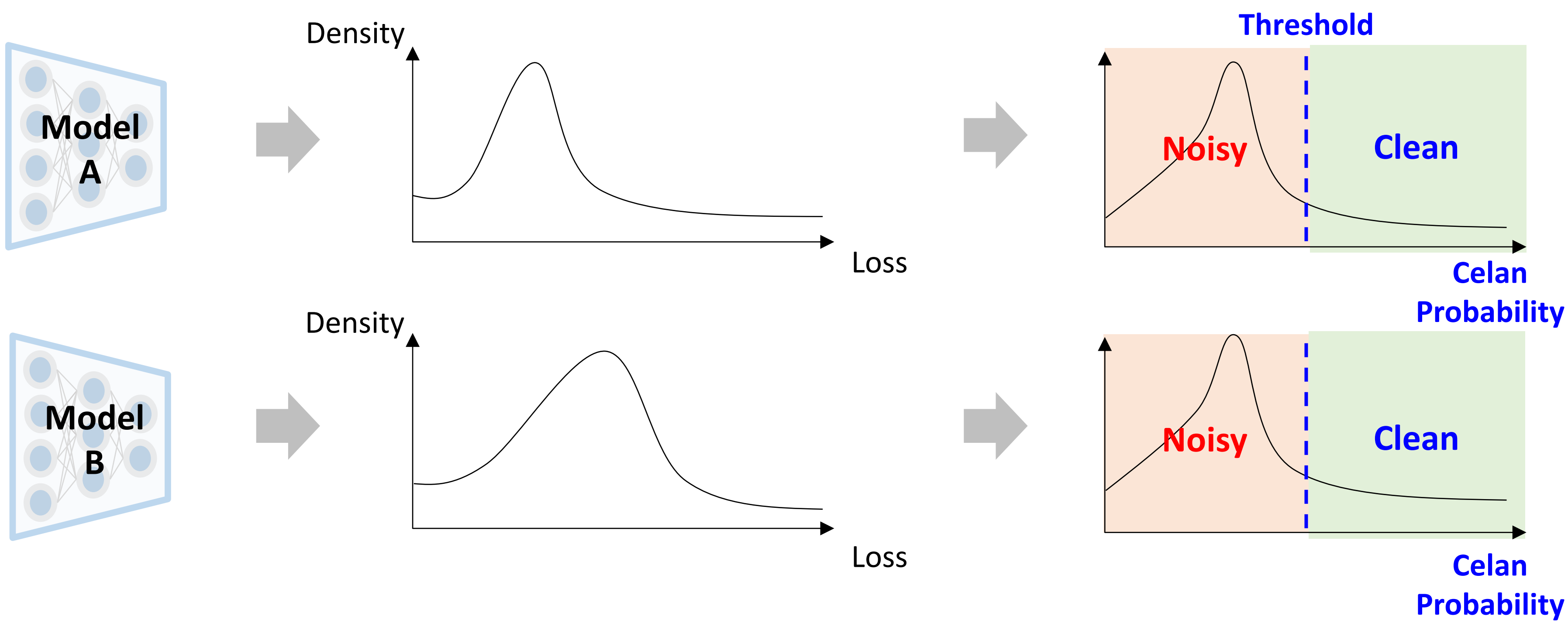
DivideMix

Method



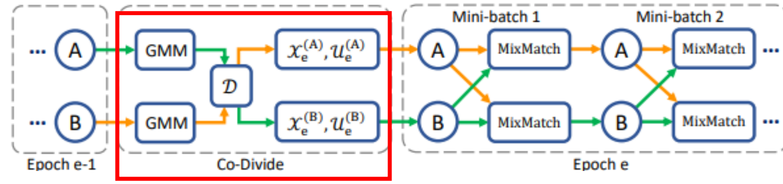
❖ Co-Divide

- 두 개의 GMM 네트워크를 사용하여 loss fitting
- 각 네트워크 별로 clean probability 산출 → **clean/noisy 데이터 구분하여 각각 labeled/unlabeled 데이터로 사용**



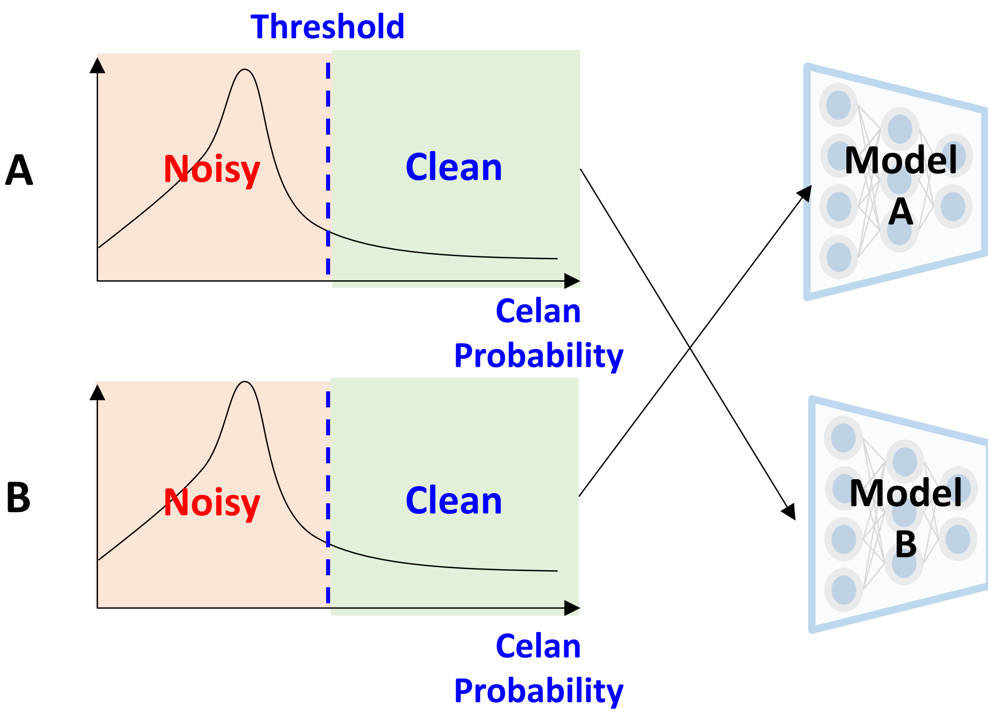
DivideMix

Method



❖ Co-Divide

- 두 개의 GMM 네트워크를 사용하여 loss fitting
- 각 네트워크 별로 clean probability 산출 → **clean/noisy 데이터 구분하여 각각 labeled/unlabeled 데이터로 사용**



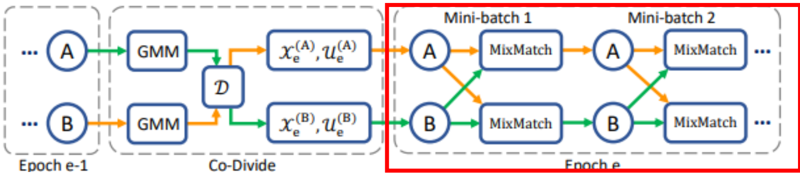
목적

Co-Divide를 통해

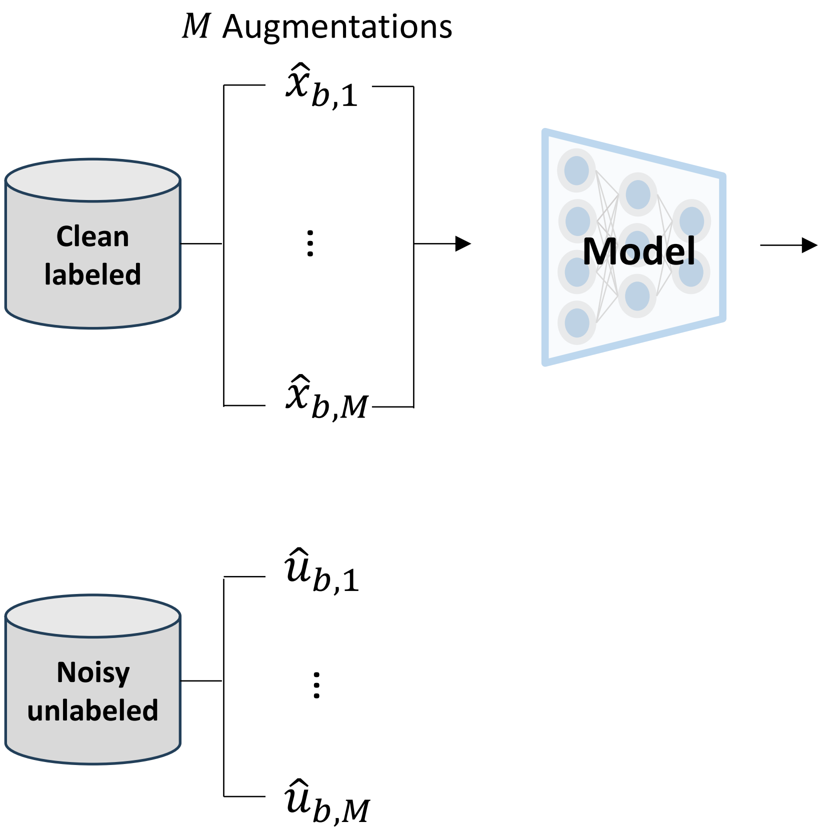
학습 과적합을 방지

DivideMix

Method



❖ Label Co-Refinement & Label Co-Guessing

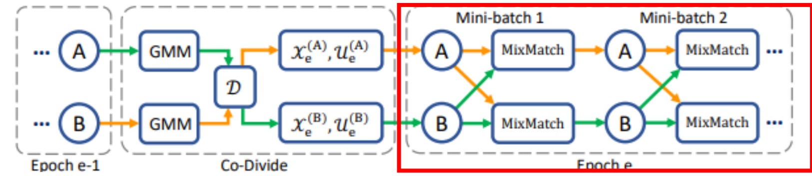


$$p_b = \frac{1}{M}\sum_m p_{model}(\hat{x}_{b,m}; \theta^{(k)})$$

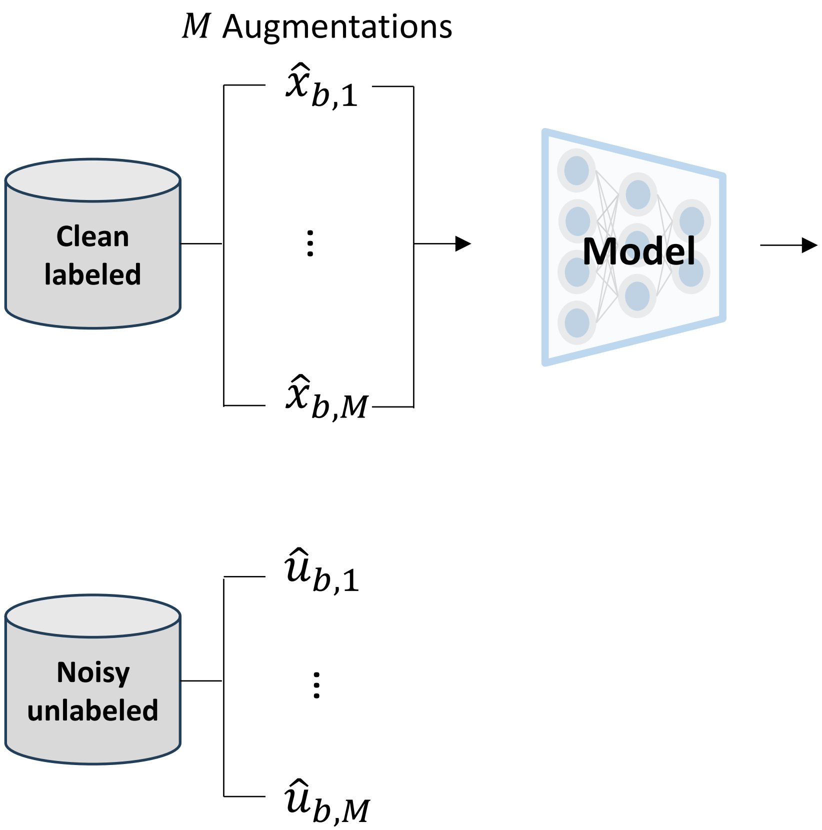
DMQA

DivideMix

Method



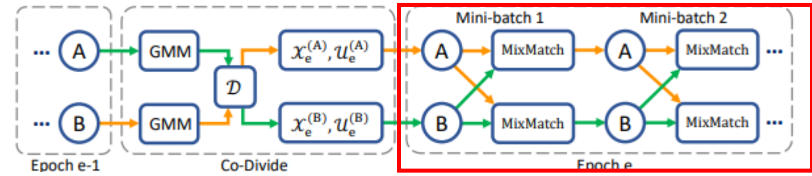
❖ Label Co-Refinement & Label Co-Guessing



$$p_b = \frac{1}{M}\sum_m p_{model}(\hat{x}_{b,m}; \theta^{(k)})$$

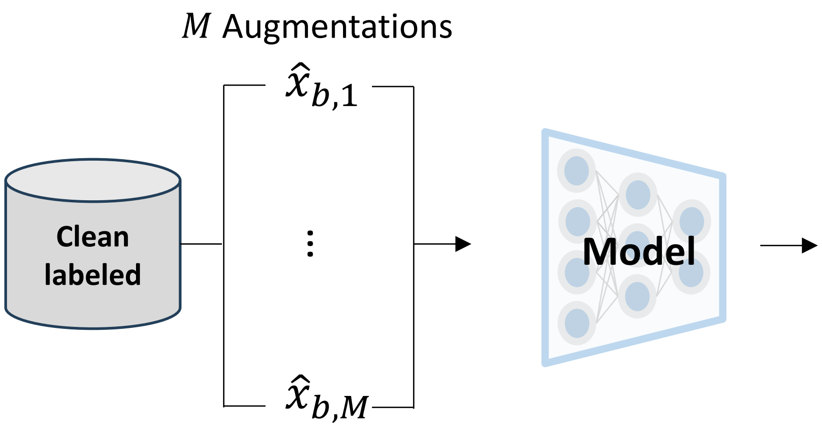
$$\bar{y}_b = \underbrace{w_b}_{\text{clean 확률}} y_b + (1 - w_b) p_b$$

DivideMix



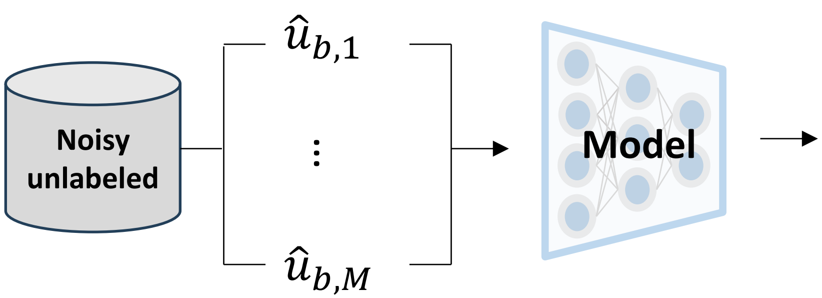
Method

❖ **Label Co-Refinement & Label Co-Guessing**



$$p_b = \frac{1}{M}\sum_m p_{model}(\hat{x}_{b,m};\theta^{(k)})$$

$$\bar{y}_b = w_b y_b + (1 - w_b) p_b$$

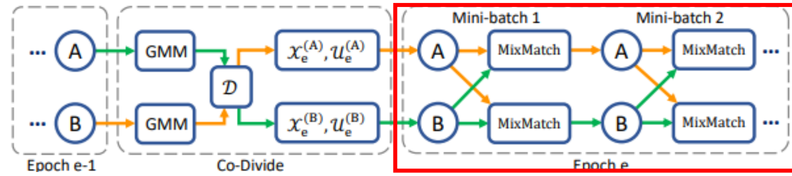


$$q_b = \frac{1}{2M}\sum_m \left(p_{model}(\hat{u}_{b,m};\theta^{(1)}) + p_{model}(\hat{u}_{b,m};\theta^{(2)})\right)$$

DMQA

DivideMix

Method

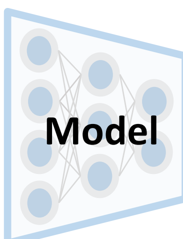


❖ MixMatch

- $L = L_x + \lambda_u L_u + \lambda_r L_{reg}$

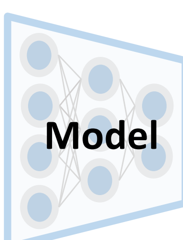
$$\hat{x}_b{}' = \lambda' \hat{x}_1{}' + (1-\lambda')\hat{x}_2{}'$$
$$\hat{y}_b{}' = \lambda' \hat{y}_1{}' + (1-\lambda')\hat{y}_2{}'$$
$$\hat{u}_b{}' = \lambda' \hat{u}_1{}' + (1-\lambda')\hat{u}_2{}'$$
$$\hat{q}_b{}' = \lambda' \hat{q}_1{}' + (1-\lambda')\hat{q}_2{}'$$

$\hat{x}_b{}' \rightarrow$



$\leftrightarrow \hat{y}_b{}'$ $\Rightarrow$ L_x

$\hat{u}_b{}' \rightarrow$



$\leftrightarrow \hat{q}_b{}'$ $\Rightarrow$ L_u

DivideMix

Experiments

❖ Main results

- Symmetric noise 세팅들에서 가장 우수한 성능 도출

Dataset		CIFAR-10				CIFAR-100			
Method/Noise ratio		20%	50%	80%	90%	20%	50%	80%	90%
Cross-Entropy	Best	86.8	79.4	62.9	42.7	62.0	46.7	19.9	10.1
	Last	82.7	57.9	26.1	16.8	61.8	37.3	8.8	3.5
Bootstrap (Reed et al., 2015)	Best	86.8	79.8	63.3	42.9	62.1	46.6	19.9	10.2
	Last	82.9	58.4	26.8	17.0	62.0	37.9	8.9	3.8
F-correction (Patrini et al., 2017)	Best	86.8	79.8	63.3	42.9	61.5	46.6	19.9	10.2
	Last	83.1	59.4	26.2	18.8	61.4	37.3	9.0	3.4
Co-teaching+* (Yu et al., 2019)	Best	89.5	85.7	67.4	47.9	65.6	51.8	27.9	13.7
	Last	88.2	84.1	45.5	30.1	64.1	45.3	15.5	8.8
Mixup (Zhang et al., 2018)	Best	95.6	87.1	71.6	52.2	67.8	57.3	30.8	14.6
	Last	92.3	77.6	46.7	43.9	66.0	46.6	17.6	8.1
P-correction* (Yi & Wu, 2019)	Best	92.4	89.1	77.5	58.9	69.4	57.5	31.1	15.3
	Last	92.0	88.7	76.5	58.2	68.1	56.4	20.7	8.8
Meta-Learning* (Li et al., 2019)	Best	92.9	89.3	77.4	58.7	68.5	59.2	42.4	19.5
	Last	92.0	88.8	76.1	58.3	67.7	58.0	40.1	14.3
M-correction (Arazo et al., 2019)	Best	94.0	92.0	86.8	69.1	73.9	66.1	48.2	24.3
	Last	93.8	91.9	86.6	68.7	73.4	65.4	47.6	20.5
DivideMix	Best	**96.1**	**94.6**	**93.2**	**76.0**	**77.3**	**74.6**	**60.2**	**31.5**
	Last	**95.7**	**94.4**	**92.9**	**75.4**	**76.9**	**74.2**	**59.6**	**31.0**

DivideMix

Experiments

❖ Main results

- Asymmetric noise 세팅에서 가장 우수한 성능 도출

Method	Best	Last
Cross-Entropy	85.0	72.3
F-correction (Patrini et al., 2017)	87.2	83.1
M-correction (Arazo et al., 2019)	87.4	86.3
Iterative-CV (Chen et al., 2019)	88.6	88.0
P-correction (Yi & Wu, 2019)	88.5	88.1
Joint-Optim (Tanaka et al., 2018)	88.9	88.4
Meta-Learning (Li et al., 2019)	89.2	88.6
DivideMix	**93.4**	**92.1**

Table 2: Comparison with state-of-the-art methods in test accuracy (%) on CIFAR-10 with 40% asymmetric noise. We re-implement all methods under the same setting.

DMQA

DivideMix

Experiments

❖ Main results

- 실제 noise 벤치마크에서 가장 우수한 성능 도출

Method	Test Accuracy
Cross-Entropy	69.21
F-correction (Patrini et al., 2017)	69.84
M-correction (Arazo et al., 2019)	71.00
Joint-Optim (Tanaka et al., 2018)	72.16
Meta-Cleaner (Zhang et al., 2019)	72.50
Meta-Learning (Li et al., 2019)	73.47
P-correction (Yi & Wu, 2019)	73.49
DivideMix	**74.76**

Table 3: Comparison with state-of-the-art methods in test accuracy (%) on Clothing1M. Results for baselines are copied from original papers.

Method	WebVision		ILSVRC12	
	top1	top5	top1	top5
F-correction (Patrini et al., 2017)	61.12	82.68	57.36	82.36
Decoupling (Malach & Shalev-Shwartz, 2017)	62.54	84.74	58.26	82.26
D2L (Ma et al., 2018)	62.68	84.00	57.80	81.36
MentorNet (Jiang et al., 2018)	63.00	81.40	57.80	79.92
Co-teaching (Han et al., 2018)	63.58	85.20	61.48	84.70
Iterative-CV (Chen et al., 2019)	65.24	85.34	61.60	84.98
DivideMix	**77.32**	**91.64**	**75.20**	**90.84**

Table 4: Comparison with state-of-the-art methods trained on (mini) WebVision dataset. Numbers denote top-1 (top-5) accuracy (%) on the WebVision validation set and the ImageNet ILSVRC12 validation set. Results for baseline methods are copied from Chen et al. (2019).

DMQA

DivideMix

Experiments

❖ Ablation study

Dataset		CIFAR-10					CIFAR-100			
Noise type		Sym.				Asym.	Sym.			
Methods/Noise ratio		20%	50%	80%	90%	40%	20%	50%	80%	90%
DivideMix	Best	**96.1**	**94.6**	**93.2**	**76.0**	**93.4**	**77.3**	**74.6**	**60.2**	**31.5**
	Last	**95.7**	**94.4**	**92.9**	**75.4**	**92.1**	**76.9**	**74.2**	**59.6**	**31.0**
DivideMix with $\theta^{(1)}$ test	Best	95.2	94.2	93.0	75.5	92.7	75.2	72.8	58.3	29.9
	Last	95.0	93.7	92.4	74.2	91.4	74.8	72.1	57.6	29.2
DivideMix w/o co-training	Best	95.0	94.0	92.6	74.3	91.9	74.8	72.3	56.7	27.7
	Last	94.8	93.3	92.2	73.2	90.6	74.1	71.7	56.3	27.2
DivideMix w/o label refinement	Best	96.0	94.6	93.0	73.7	87.7	76.9	74.2	58.7	26.9
	Last	95.5	94.2	92.7	73.0	86.3	76.4	73.9	58.2	26.3
DivideMix w/o augmentation	Best	95.3	94.1	92.2	73.9	89.5	76.5	73.1	58.2	26.9
	Last	94.9	93.5	91.8	73.0	88.4	76.2	72.6	58.0	26.4
Divide and MixMatch	Best	94.1	92.8	89.7	70.1	86.5	73.7	70.5	55.3	25.0
	Last	93.5	92.3	89.1	68.6	85.2	72.4	69.7	53.9	23.7

Table 5: Ablation study results in terms of test accuracy (%) on CIFAR-10 and CIFAR-100.

DMQA

Bayesian DivideMix++

Paper

❖ Bayesian DivideMix++ for Enhanced Learning with Noisy Labels (2024, Neural Networks)

- **불확실성**을 기반으로 DivideMix 보완

Bayesian DivideMix++ for Enhanced Learning with Noisy Labels

Bhalaji Nagarajan [a,*], Ricardo Marques [a,c], Eduardo Aguilar [a,b,c], Petia Radeva [a,c]

[a] *Dept. de Matemàtiques i Informàtica, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007, Barcelona, Spain*

[b] *Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Avenida Angamos 0610, 1270709, Antofagasta, Chile*

[c] *Computer Vision Center, Cerdanyola (Barcelona), Spain*

ARTICLE INFO

Keywords:
Learning with noisy labels
Neural network memorization
Data augmentation
Self-supervised pre-training
Label uncertainty
Monte-Carlo dropouts

ABSTRACT

Leveraging inexpensive and human intervention-based annotating methodologies, such as crowdsourcing and web crawling, often leads to datasets with noisy labels. Noisy labels can have a detrimental impact on the performance and generalization of deep neural networks. Robust models that are able to handle and mitigate the effect of these noisy labels are thus essential. In this work, we explore the open challenges of neural network memorization and uncertainty in creating robust learning algorithms with noisy labels. To overcome them, we propose a novel framework called "Bayesian DivideMix++" with two critical components: (i) DivideMix++, to enhance the robustness against memorization and (ii) Monte-Carlo MixMatch, which focuses on improving the effectiveness towards label uncertainty. DivideMix++ improves the pipeline by integrating the warm-up and augmentation pipeline with self-supervised pre-training and dedicated different data augmentations for loss analysis and backpropagation. Monte-Carlo MixMatch leverages uncertainty measurements to mitigate the influence of uncertain samples by reducing their weight in the data augmentation MixMatch step. We validate our proposed pipeline using four datasets encompassing various synthetic and real-world noise settings. We demonstrate the effectiveness and merits of our proposed pipeline using extensive experiments. Bayesian DivideMix++ outperforms the state-of-the-art models by considerable differences in all experiments. Our findings underscore the potential of leveraging these modifications to enhance the performance and generalization of deep neural networks in practical scenarios.

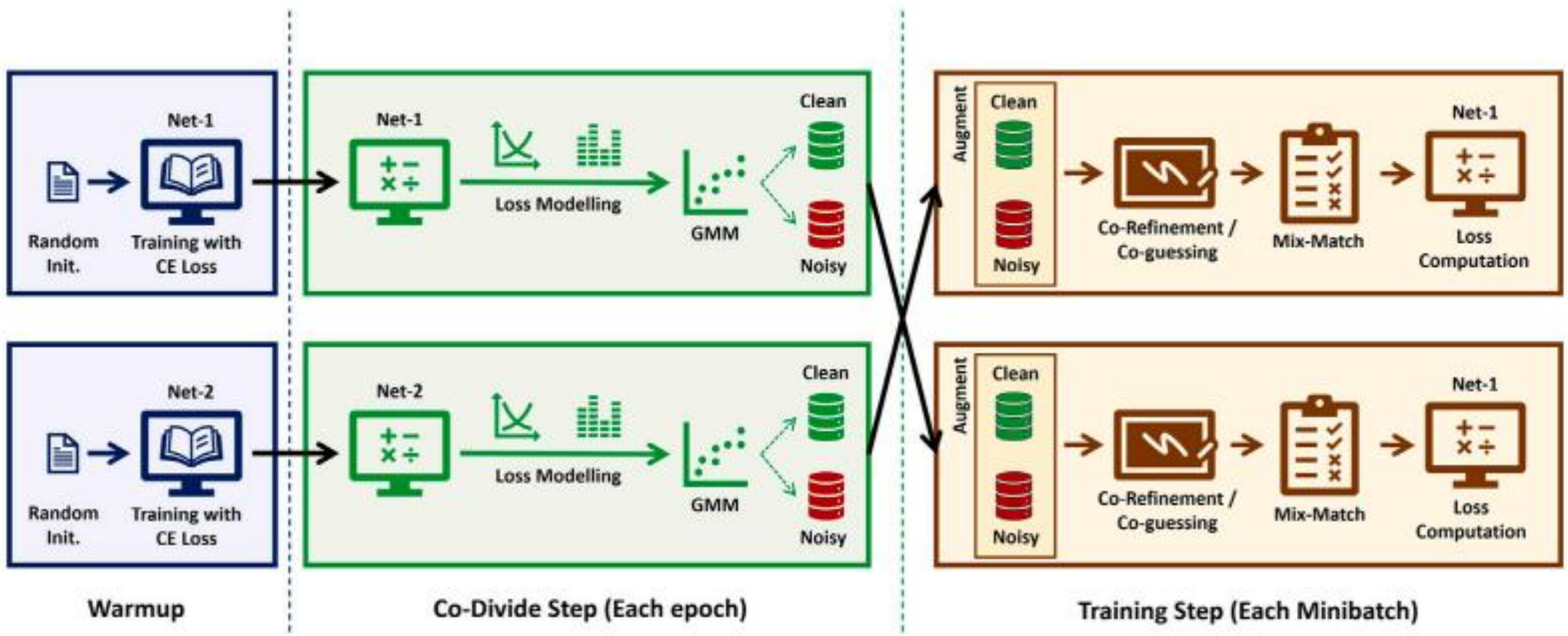
Nagarajan, B., Marques, R., Aguilar, E., & Radeva, P. (2024). Bayesian DivideMix++ for enhanced learning with noisy labels. *Neural Networks*, *172*, 106122.

Bayesian DivideMix++

Method

❖ DivideMix의 문제점 개선

1. Model confidence를 예측 확률 분포가 아닌 **불확실성**을 사용하자
2. Supervised pre-training은 noisy label 상황에서 악영향을 미치기 때문에 **unsupervised pre-training**을 통해 사전에 noisy와 clean을 잘 구분할 수 있는 능력을 기르자
3. **Strong augmentations 기법**들을 추가하여 noisy label 상황에서 모델을 보다 강건하게 학습하자



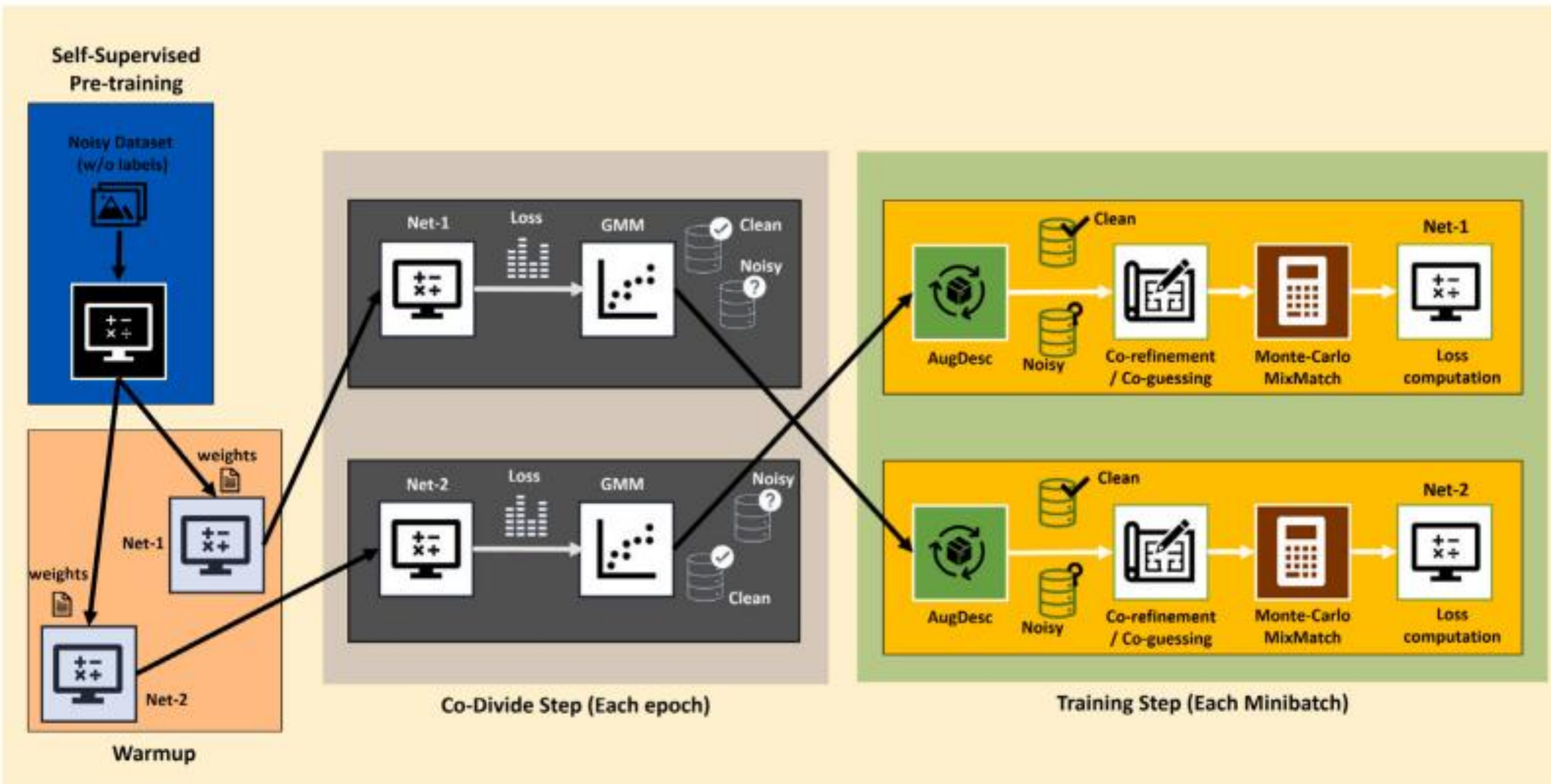
[DivideMix]

DMQA

Bayesian DivideMix++

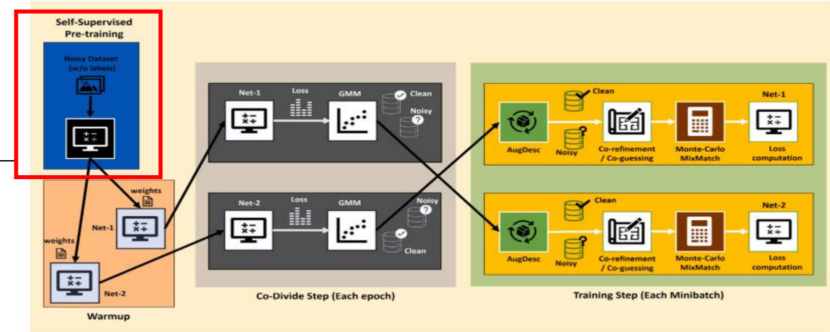
Method

❖ Overall Process



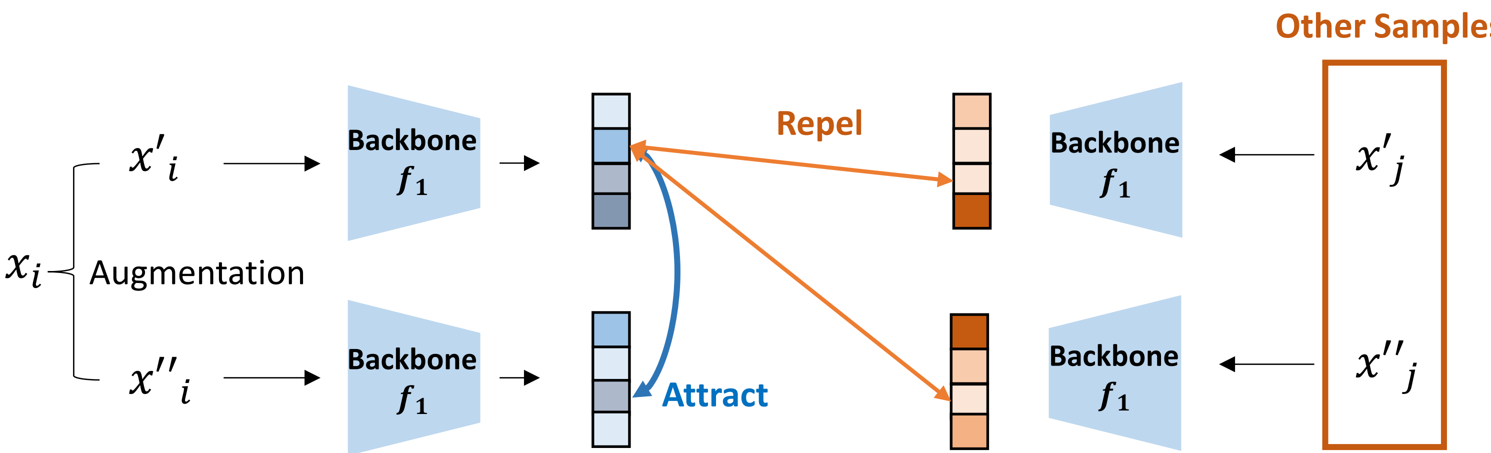
Bayesian DivideMix++

Method



❖ **Self-supervised pre-training**

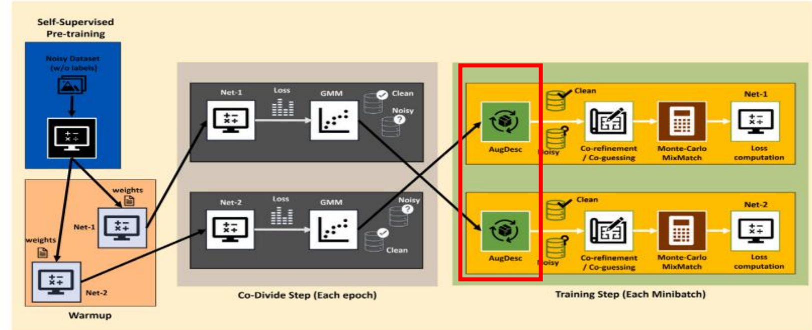
- Noisy label의 악영향을 최소화 하고자 함
- 모델 파라미터를 random initialize하는 것이 아닌 사전 학습된 **SimCLR** 파라미터로 사용



Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PmLR.

DMQA

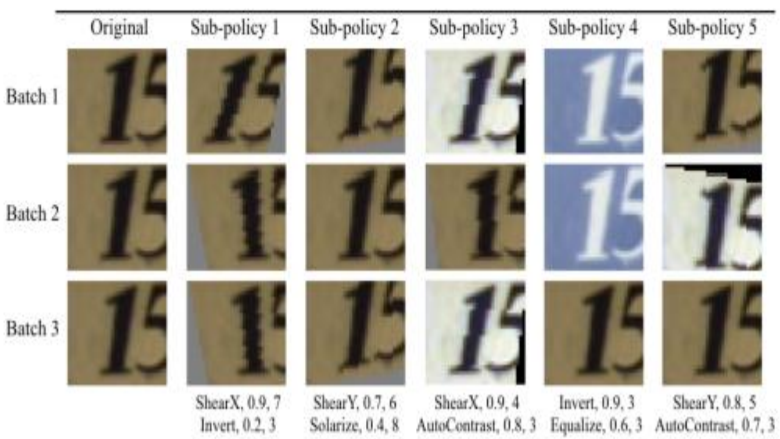
Bayesian DivideMix++



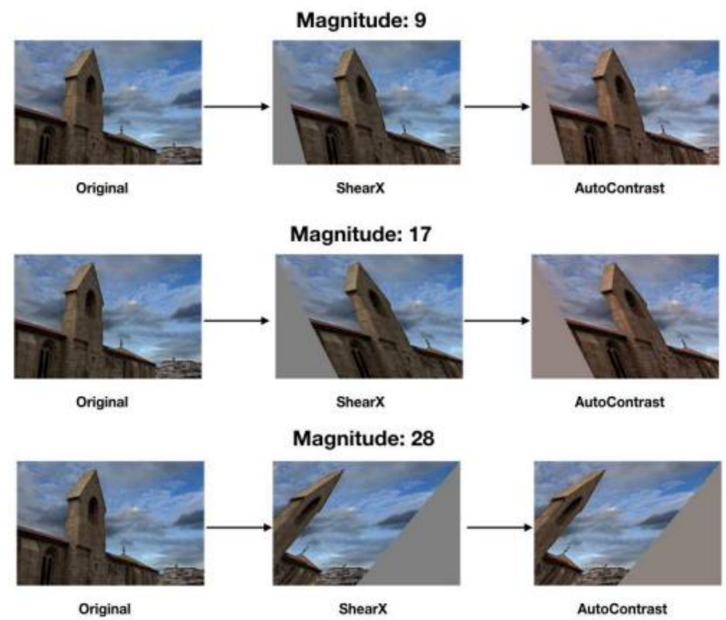
Method

❖ Incorporating strong augmentation

- High noise 세팅에서 강건하도록 strong augmentation 기법 추가
  - ➢ AutoAugment / RandAugment



[AutoAugment]



[RandAugment]

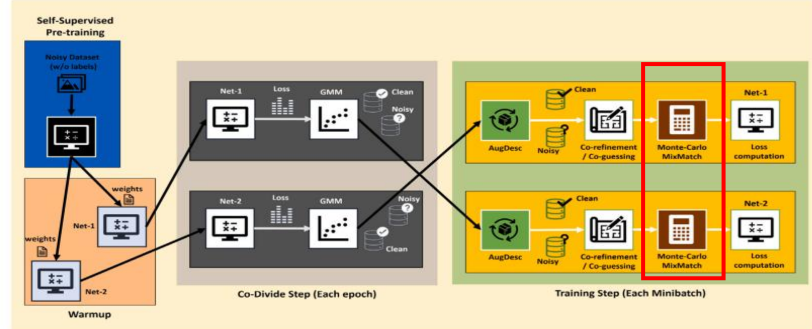
Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (1805). Autoaugment: Learning augmentation policies from data. arXiv 2018. *arXiv preprint arXiv:1805.09501*, *2*.

Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 702-703).

DMQA

Bayesian DivideMix++

Method



❖ Monte-Carlo MixMatch (MC-MixMatch)

- Mixup 단계에서 불확실성이 높은 샘플에 집중하여 모델이 noise에 더 강건해지도록 함
- $H(x_1)$: Mixup 적용 시 첫 번째 샘플의 불확실성

$\boldsymbol{\lambda_{unc}} \times$

x_1

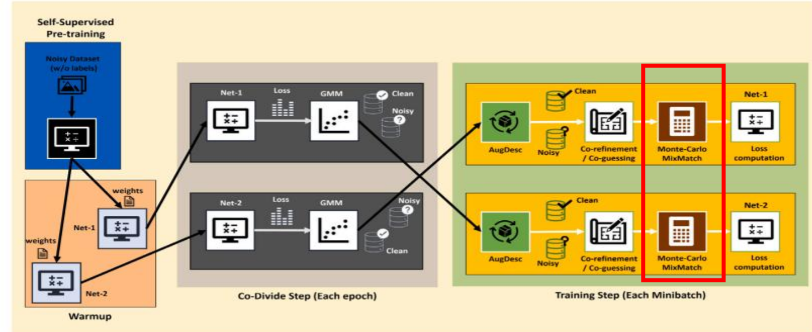
$(\mathbf{1} - \boldsymbol{\lambda_{unc}}) \times$

x_2

$$\lambda \sim Beta(\alpha, \alpha)$$
$$\lambda' = \max(\lambda, 1 - \lambda)$$
$$\lambda_{unc} = 0.5 \times (H(x_1) \times \lambda' + 1)$$
$$x' = \lambda_{unc} x_1 + (1 - \lambda_{unc}) x_2$$
$$y' = \lambda_{unc} y_1 + (1 - \lambda_{unc}) y_2$$

Bayesian DivideMix++

Method



❖ Monte-Carlo MixMatch (MC-MixMatch)

- Mixup 단계에서 불확실성이 높은 샘플에 집중하여 모델이 noise에 더 강건해지도록 함
- $H(x_1)$: Mixup 적용 시 첫 번째 샘플의 불확실성

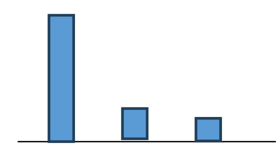
$\boldsymbol{\lambda_{unc}} \times$

x_1

$(\mathbf{1} - \boldsymbol{\lambda_{unc}}) \times$

x_2

$$\lambda \sim Beta(\alpha, \alpha)$$
$$\lambda' = \max(\lambda, 1 - \lambda)$$
$$\lambda_{unc} = 0.5 \times (H(x_1) \times \lambda' + 1)$$
$$x' = \lambda_{unc} x_1 + (1 - \lambda_{unc}) x_2$$
$$y' = \lambda_{unc} y_1 + (1 - \lambda_{unc}) y_2$$



확률 분포

$$\boldsymbol{H(x_i) = -\sum_{c=1}^{C} p_c(x_i) \log(p_c(x_i))}$$

DMQA

Bayesian DivideMix++

Experiments

❖ Main results

- Symmetric noise 세팅에서 대부분 가장 우수한 성능 도출

Table 3

CIFAR-10 — Performance comparison (test accuracy (%, mean ± std over five runs)) with SoTA methods on different symmetric noise settings.

Method / Noise ratio		20%	50%	80%	90%
DivideMix (ICLR, 2020)	Best	96.1	94.6	93.2	76.0
Li et al. (2020)	Last	95.7	94.4	92.9	75.4
DM + AugDesc (CVPR, 2021)	Best	96.3	95.4	93.8	91.9
Nishi et al. (2021)	Last	96.2	95.1	93.6	91.8
DM + C2D (WACV, 2022)	Best	96.43±0.07	95.32±0.12	94.40±0.04	93.57±0.09
Zheltonozhskii et al. (2022)	Last	96.23±0.09	95.15±0.16	94.30±0.12	93.42±0.09
Sel-CL+ (CVPR, 2022)	Best	95.5	93.9	89.2	81.9
Li et al. (2022)	Last	–	–	–	–
UNICON (CVPR, 2022)	Best	96.0	95.6	93.9	90.8
Karim et al. (2022)	Last	–	–	–	–
ULC (AAAI, 2022)	Best	96.1	95.2	94.0	86.4
Huang, Bai, et al. (2022)	Last	95.9	94.7	93.2	85.8
LongReMix (PR, 2023)	Best	96.3±0.1	95.1±0.1	93.8±0.2	79.9±2.7
Cordeiro, Sachdeva, Belagiannis, Reid, and Carneiro (2023)	Last	96.0±0.1	94.8±0.1	93.8±0.2	79.1±3.1
Lipschitz Reg. (PR, 2023)	Best	96.2	95.2	93.4	85.0
Miao, Wu, Xu, Zuo, and Meng (2023)	Last	95.7	94.8	93.1	84.3
Bayesian DivideMix++	Best	96.39±0.06	95.68±0.09	95.25±0.08	94.46±0.15
	Last	96.13±0.07	95.40±0.11	94.97±0.02	94.20±0.12

Bayesian DivideMix++

Experiments

❖ Main results

- Symmetric noise 세팅에서 대부분 가장 우수한 성능 도출

Table 4
CIFAR-100 — Performance comparison (test accuracy (%, mean ± std over five runs)) with SoTA methods on different symmetric noise settings.

Method / Noise ratio		20%	50%	80%	90%
DivideMix (ICLR, 2020)	Best	77.3	74.6	60.2	31.5
Li et al. (2020)	Last	76.9	74.2	59.6	31.0
DM + AugDesc (CVPR, 2021)	Best	79.5	77.2	66.4	41.2
Nishi et al. (2021)	Last	79.2	77.0	66.1	40.9
DM + C2D (WACV, 2022)	Best	78.69±0.17	76.43±0.25	67.78±0.30	58.7±0.31
Zheltonozhskii et al. (2022)	Last	78.32±0.35	76.07±0.41	67.43±0.30	58.45±0.30
Sel-CL+ (CVPR, 2022)	Best	76.5	72.4	59.6	48.8
Li et al. (2022)	Last	–	–	–	–
UNICON (CVPR, 2022)	Best	78.9	77.6	63.9	44.8
Karim et al. (2022)	Last	–	–	–	–
ULC (AAAI, 2022)	Best	77.3	74.9	61.2	34.5
Huang, Bai, et al. (2022)	Last	77.1	74.3	60.8	34.1
LongReMix (PR, 2023)	Best	77.9±0.2	75.5±0.2	62.3±0.5	34.7±0.3
Cordeiro et al. (2023)	Last	77.5±0.2	74.9±0.2	61.7±0.5	30.7±5.9
Lipschitz Reg. (PR, 2023)	Best	78.8	75.3	62.6	37.5
Miao et al. (2023)	Last	77.4	74.3	61.0	34.7
Bayesian DivideMix++	Best	**80.02±0.03**	**78.31±0.14**	**70.01±0.23**	**61.15±0.34**
	Last	79.56±0.13	77.71±0.13	69.55±0.22	60.70±0.42

DMQA

Bayesian DivideMix++

❖ Main results

- Asymmetric noise 세팅에서 가장 우수한 성능 도출

Table 5
CIFAR-100 — Asymmetric 40% Noise Performance comparison (test accuracy (%, mean ± std over five runs)) with SoTA methods.

Method		Test accuracy
DM + C2D (WACV, 2022)	Best	75.48±0.16
Zheltonozhskii et al. (2022)	Last	75.06±0.16
Sel-CL+ (CVPR, 2022)	Best	74.20
Li et al. (2022)	Last	–
UNICON (CVPR, 2022)	Best	74.80
Karim et al. (2022)	Last	–
LongReMix (PR, 2023)	Best	59.80±0.1
Cordeiro et al. (2023)	Last	54.90±0.4
Bayesian DivideMix++	Best	76.52±0.12
	Last	76.06±0.13

DMQA

Bayesian DivideMix++

Experiments

❖ Main results

- Top-1 accuracy 기준 제일 좋거나 동등한 성능을 보임

Table 6
WebVision — Performance comparison (test accuracy (%, mean ± std over five runs)) with SoTA methods. Results are reported for both WebVision and ILSVRC12 (ImageNet) validation sets. († - *results reported in DM + C2D (Zheltonozhskii et al., 2022). DivideMix results are reported only on InceptionResNet-V2).*

Validation Set / Method	WebVision Top-1	WebVision Top-5	ILSVRC12 Top-1	ILSVRC12 Top-5
DivideMix (ICLR, 2020) † Li et al. (2020)	76.32±0.36	90.65±0.16	74.42±0.29	91.21±0.12
DM + C2D (WACV, 2022) Zheltonozhskii et al. (2022)	79.42±0.34	92.32±0.33	78.57±0.37	93.04±0.10
UNICON (CVPR, 2022) Karim et al. (2022)	77.60	93.44	75.29	93.72
Bayesian DivideMix++	80.12±0.28	92.40±0.30	78.51±0.28	92.67±0.42

Table 7
Clothing1M — Performance comparison (test accuracy (%, mean ± std over five runs)) with SoTA methods.

Method	Test accuracy
DivideMix (ICLR, 2020) Li et al. (2020)	74.38[a]
DM + AugDesc (CVPR, 2021) Nishi et al. (2021)	74.57[a]
DM + C2D (WACV, 2022) Zheltonozhskii et al. (2022)	73.96[a]
ULC (AAAI, 2022) Huang, Bai, et al. (2022)	74.9±0.2
LongReMix (PR, 2023) Cordeiro et al. (2023)	74.38
Lipschitz Reg. (PR, 2023) Miao et al. (2023)	74.87
DivideMix++ (Deterministic)	74.09
DivideMix++ (with Dropouts)	74.13
Bayesian DivideMix++	74.81±0.09

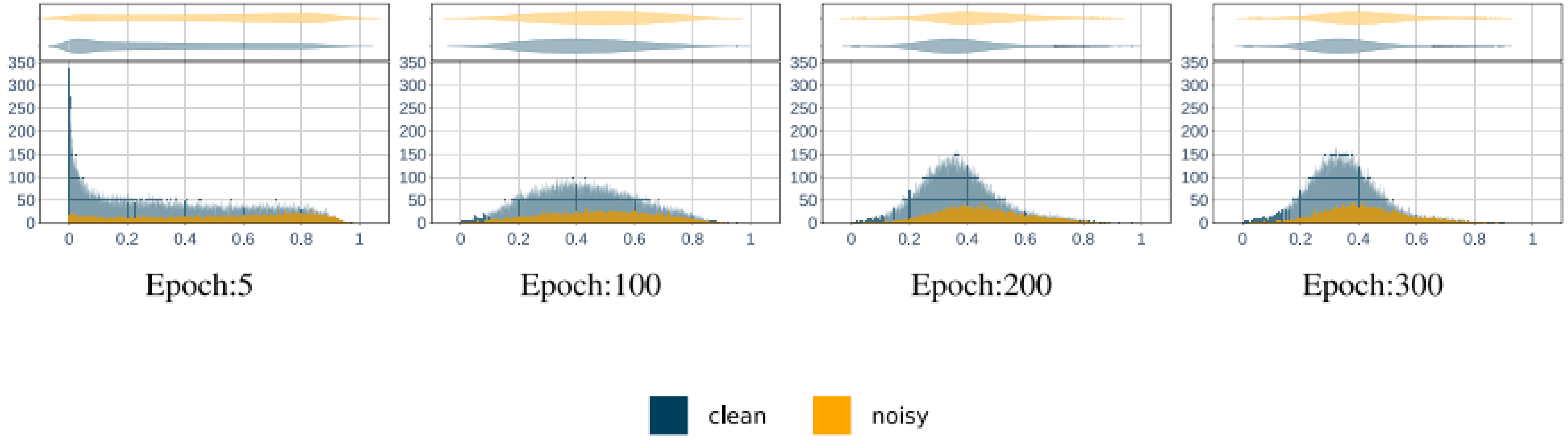
DMQA

Bayesian DivideMix++

Experiments

❖ 불확실성 시각화

- 학습이 진행될수록 noisy 데이터가 비교적 큰 불확실성 분포를 가짐



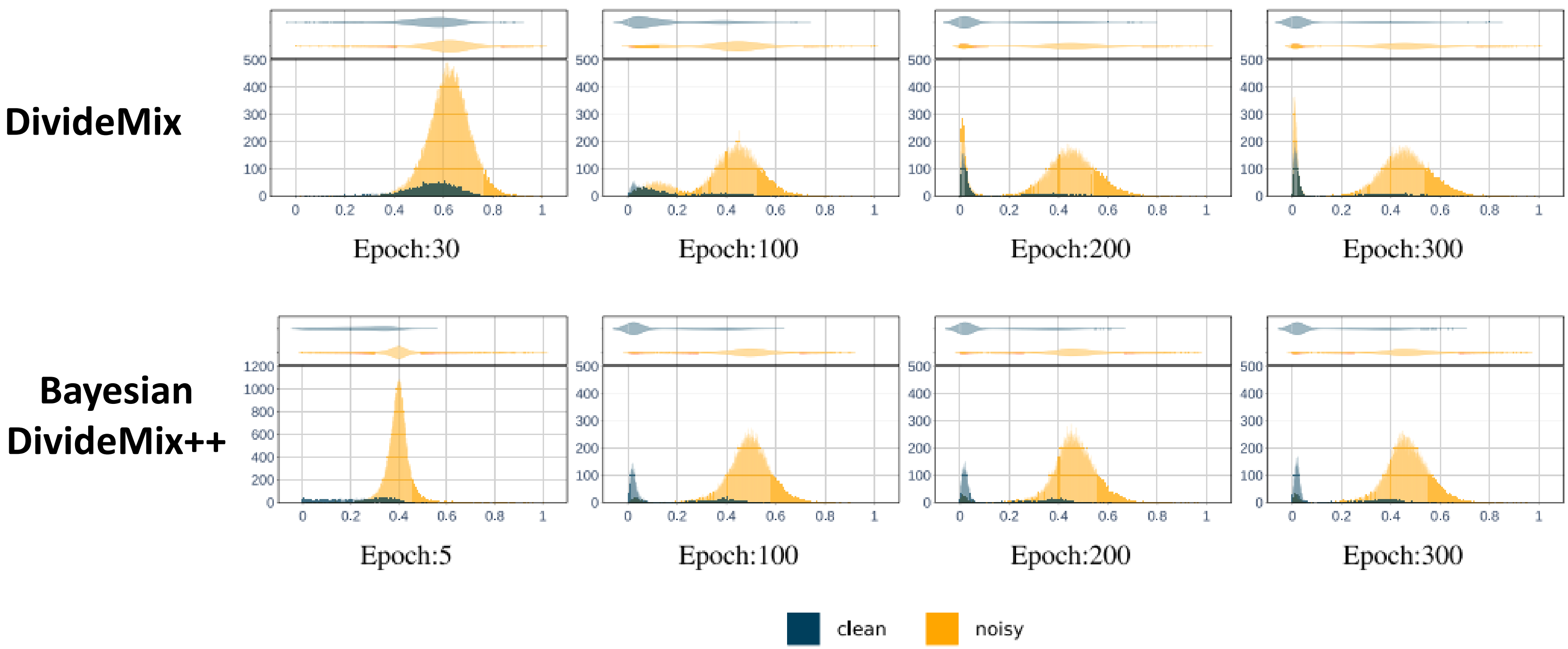
DMQA

Bayesian DivideMix++

Experiments

❖ 학습 loss 시각화

- 학습이 진행될수록 noisy와 clean 데이터 사이의 loss 분포 차이가 심해짐
  - ➢ **효과적으로 clean labeled / noisy unlabeled로 구분하여 학습 가능**



DMQA

Bayesian DivideMix++

Experiments

❖ Dropout position 분석

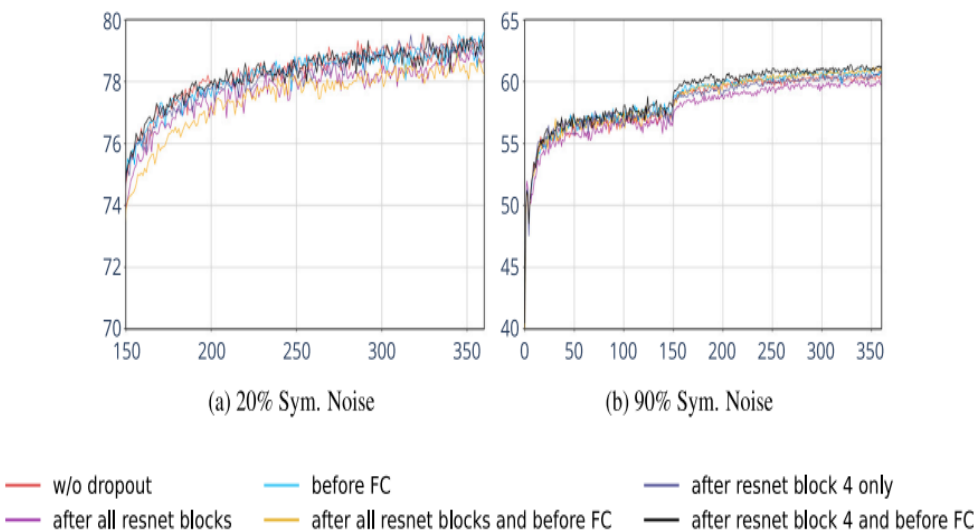


Fig. 8. Dropout position. Different configurations to select a better dropout position.

Table 11
Position of MC-Dropout. Difference in test accuracy (%) on CIFAR-100 (20% and 90% symmetric noise) with different combination of dropout layers. The dropout ratio was set to 0.1 in all cases.

Position of MC-Dropout		20%	90%
After all blocks	Best	79.05	60.15
	Last	78.77	59.86
Only after block 4	Best	79.51	60.85
	Last	79.13	60.61
Only before FC	Best	79.48	61.09
	Last	78.89	60.62
After all blocks & before FC	Best	79.00	61.18
	Last	78.44	60.99
After block 4 & before FC	Best	79.48	61.39
	Last	79.16	61.13

Manifold DivideMix

Paper

- **Manifold DivideMix: A Semi-Supervised Contrastive Learning Framework for Severe Label Noise (2024, CVPR Workshop)**
  - Out-Of-Distribution noise까지 추가된 문제 상황

Manifold DivideMix: A Semi-Supervised Contrastive Learning Framework for Severe Label Noise

Fahimeh Fooladgar[1] Minh Nguyen Nhat To[1] Parvin Mousavi[2] Purang Abolmaesumi[1]
[1] Department of Electrical and Computer Engineering, University of British Columbia, CA
[2] School of Computing, Queen's University, CA

fahimeh.fooladgar@ubc.ca, mtrcl@student.ubc.ca, mousavi@queensu.ca, purang@ece.ubc.ca

Abstract

Deep neural networks have proven to be highly effective when large amounts of data with clean labels are available. However, their performance degrades when training data contains noisy labels, leading to poor generalization on the test set. Real-world datasets contain noisy label samples that either have similar visual semantics to other classes (in-distribution) or have no semantic relevance to any class (out-of-distribution) in the dataset. Most state-of-the-art methods leverage ID labeled noisy samples as unlabeled data for semi-supervised learning, but OOD labeled noisy samples cannot be used in this way because they do not belong to any class within the dataset. Hence, in this paper, we propose incorporating the information from all the training data by leveraging the benefits of self-supervised training. Our method aims to extract a meaningful and generalizable embedding space for each sample regardless of its label. Then, we employ a simple yet effective K-nearest neighbor method to remove portions of out-of-distribution samples. By discarding these samples, we propose an iterative "Manifold DivideMix" algorithm to find clean and noisy samples, and train our model in a semi-supervised way. In addition, we propose "MixEMatch", a new algorithm for the semi-supervised step that involves mixup augmentation at the input and final hidden representations of the model. This will extract better representations by interpolating both in the input and manifold spaces. Extensive experiments on multiple synthetic-noise image benchmarks and real-world web-crawled datasets demonstrate the effectiveness of our proposed framework. Code is available at https://github.com/Fahim-F/ManifoldDivideMix.

amount of clean, annotated data [13]. However, their primary weakness is also the vast number of clean labeled examples needed for training. Collecting and manually annotating such data could be complex, time-consuming, and costly, particularly in certain fields. Meanwhile, open-source online data that can be automatically annotated using search engine queries and user tags is the backbone of most large-scale data collection methods [22, 37]. However, this annotation approach will certainly result in label noise. Therefore, it is challenging to train DNNs using such data, as they can effectively memorize noisy random labels during training [4, 23].

Several methods have been developed to deal with label noise in automatically annotated datasets, such as semi-supervised learning [5, 31], self-supervised learning [6], and robust training [19]. These methods can be classified into two primary categories. The first category assumes that the true labels of noisy samples are included in the label set (i.e. in-distribution labeled noisy samples (ID samples). The community has invested a lot of effort in designing robust methods to train DNNs in the presence of ID label noise [11, 19]. The second category arises from the observation that techniques that are robust to ID noise tend to perform poorly when applied in more realistic settings (with real-world label noise). In fact, the authors of [2] suggest that most of the label noise in web-crawled datasets is out-of-distribution (OOD) label noise, which means that the true labels for noisy samples cannot be inferred from the distribution (we call them OOD samples). To evaluate this, they randomly collected three small but representative sample sets from the WebVision 1.0 dataset [21] to determine the typical level of noise present in web-crawled, automatically annotated datasets. They reported that approximately 70%

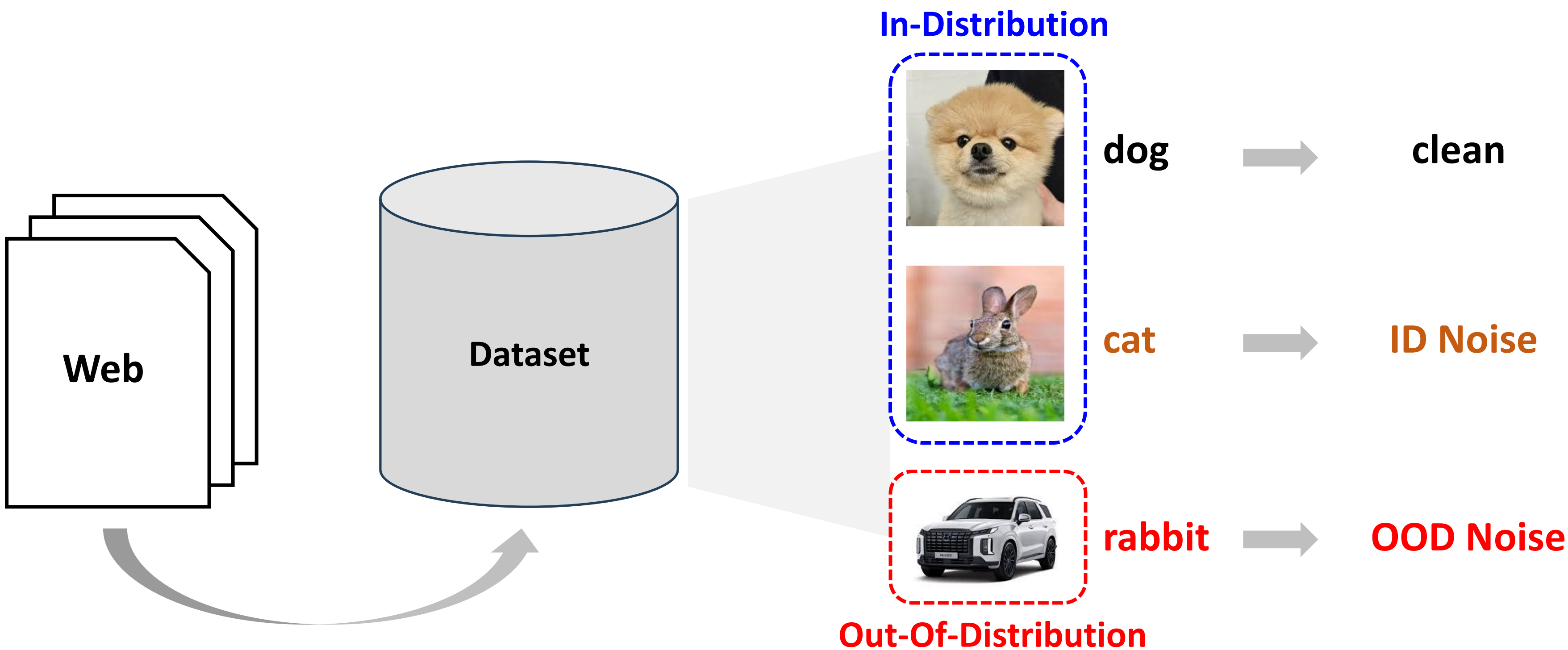
Fooladgar, F., To, M. N. N., Mousavi, P., & Abolmaesumi, P. (2024). Manifold DivideMix: A semi-supervised contrastive learning framework for severe label noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4012-4021).

Manifold DivideMix

Background

❖ **Semi-Supervised Learning for Label Noise Learning with Out-Of-Distribution Noise**

- Web을 통해 이미지 데이터셋을 수집하다 보면, **분포가 아예 다른 종류의 데이터**가 같이 수집될 수 있음
- 학습 데이터에 Out-Of-Distribution(OOD) noise와 In-Distribution(ID) noise가 혼재함을 가정

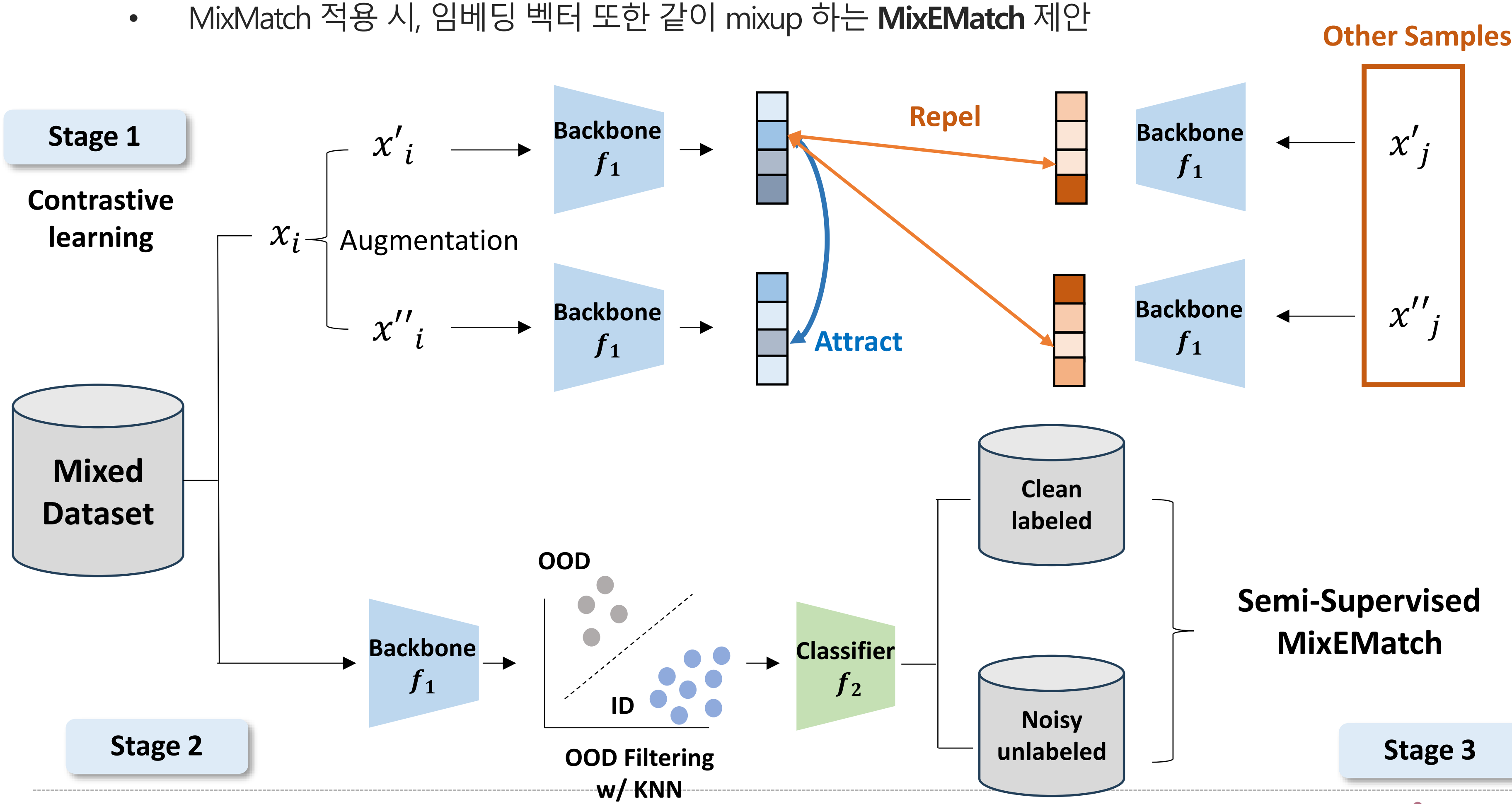


Manifold DivideMix

Method

❖ Overall Process

- **Contrastive learning** 기반 사전학습을 통해 **OOD와 ID 샘플을 잘 구분**
- MixMatch 적용 시, 임베딩 벡터 또한 같이 mixup 하는 **MixEMatch** 제안



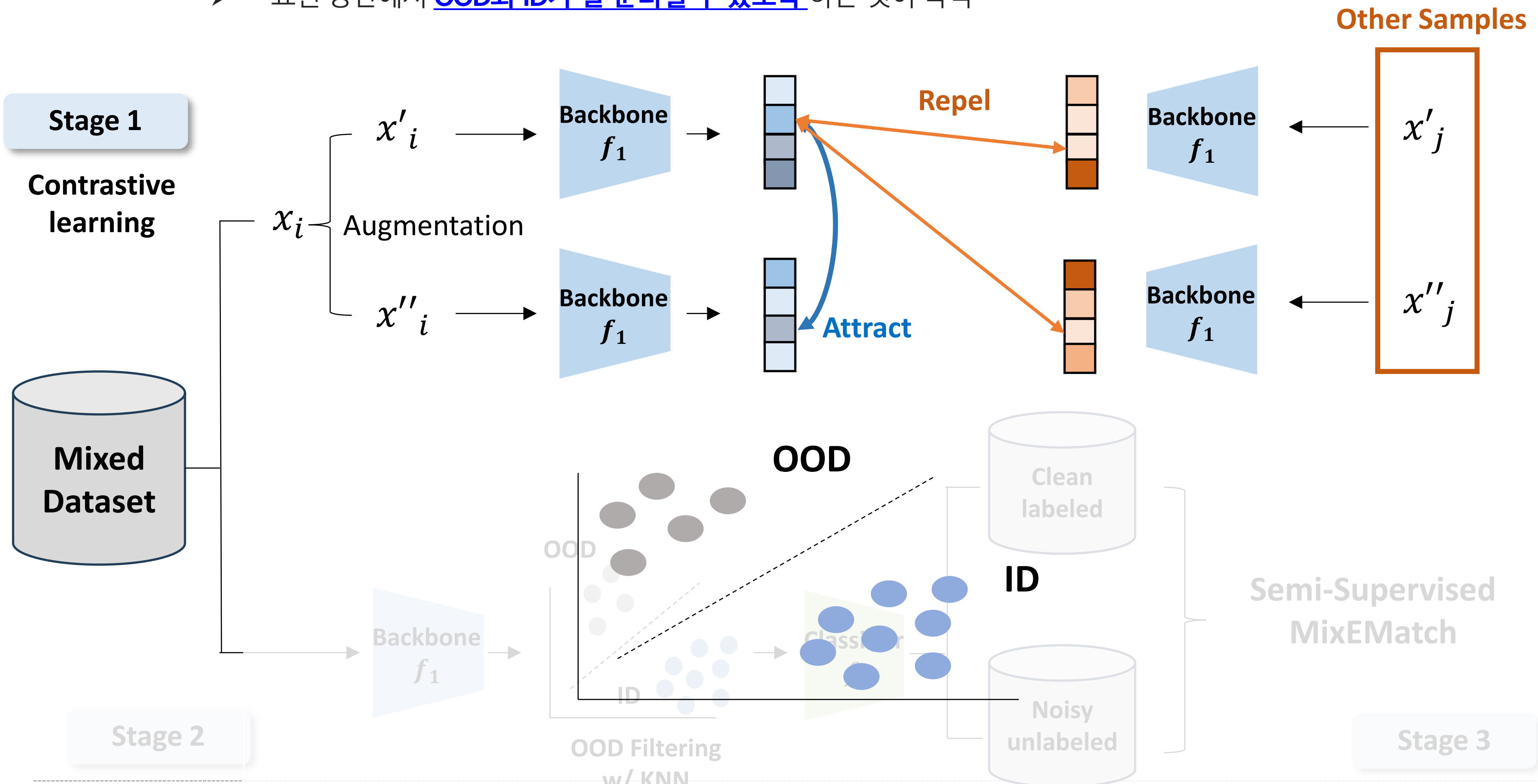
DMQA

Manifold DivideMix

Method

❖ Pre-training with contrastive learning

- Unsupervised contrastive learning을 통해 올바른 표현 학습 수행
  - ➢ 표현 공간에서 **OOD와 ID가 잘 분리될 수 있도록** 하는 것이 목적



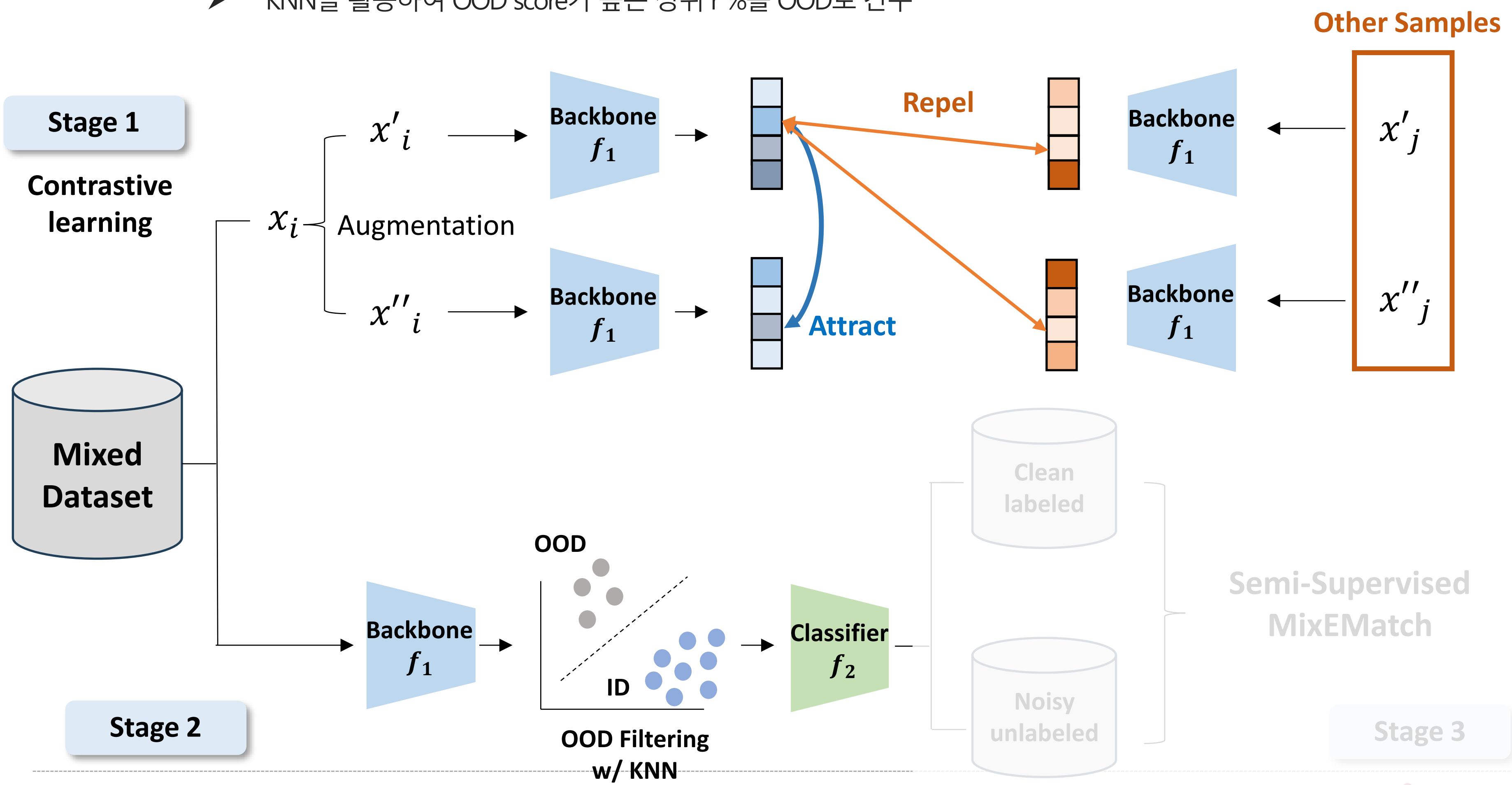
DMQA

Manifold DivideMix

Method

❖ OOD noise filtering

- 사전학습 된 백본 네트워크를 통해 표현 공간에서 OOD와 ID 구분
  - ➢ KNN을 활용하여 OOD score가 높은 상위 r %를 OOD로 간주



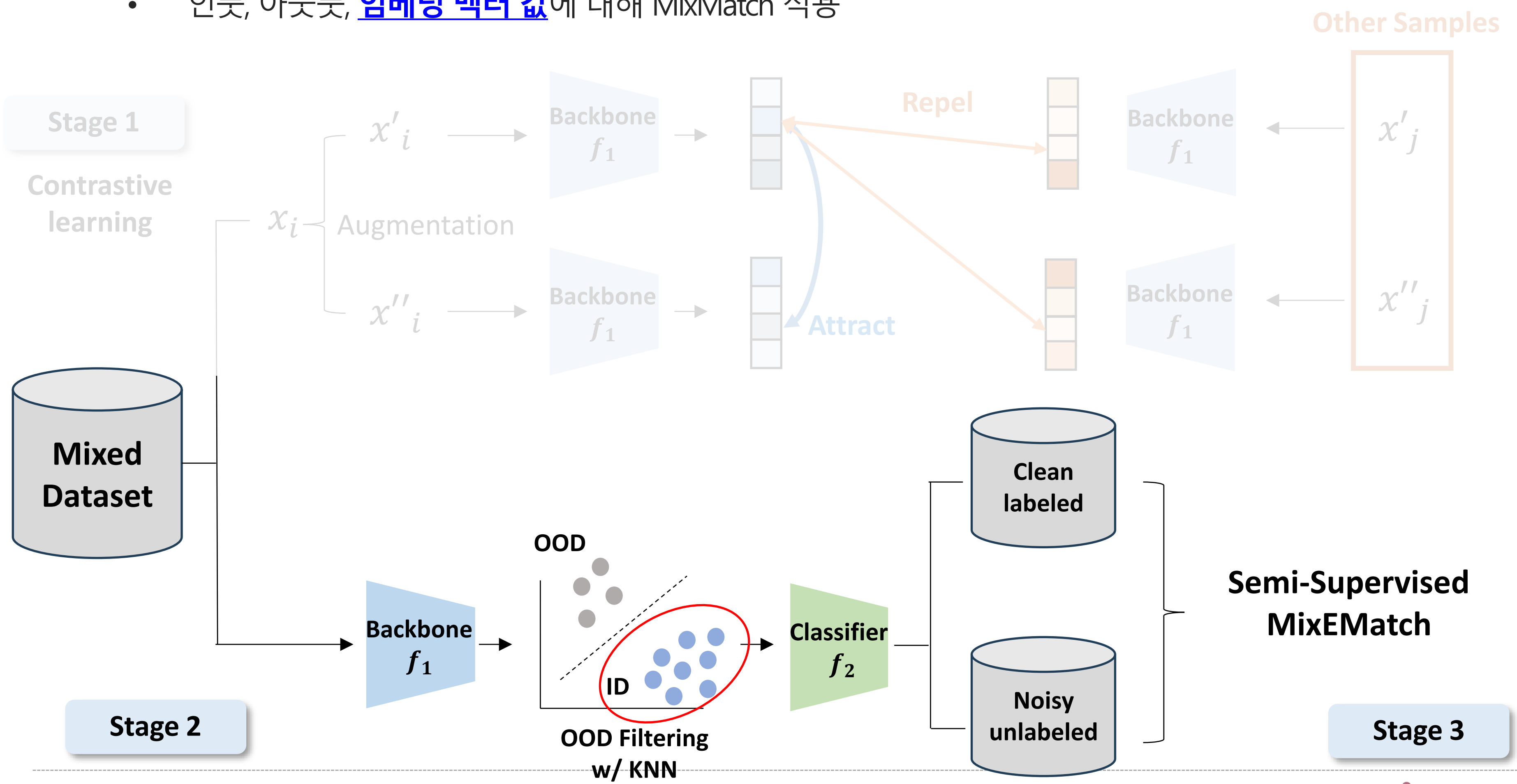
DMQA

Manifold DivideMix

Method

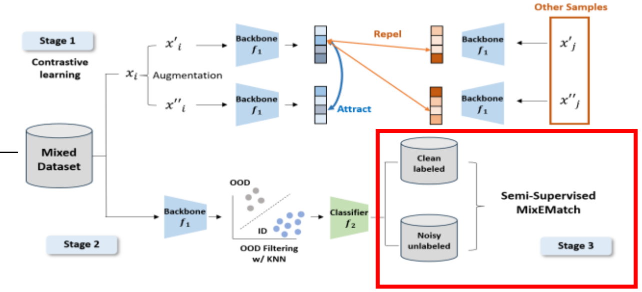
❖ Semi-Supervised Learning with MixEMatch

- 선정된 ID 데이터를 loss 기준으로 clean / noise 구분
- 인풋, 아웃풋, **임베딩 벡터 값**에 대해 MixMatch 적용



DMQA

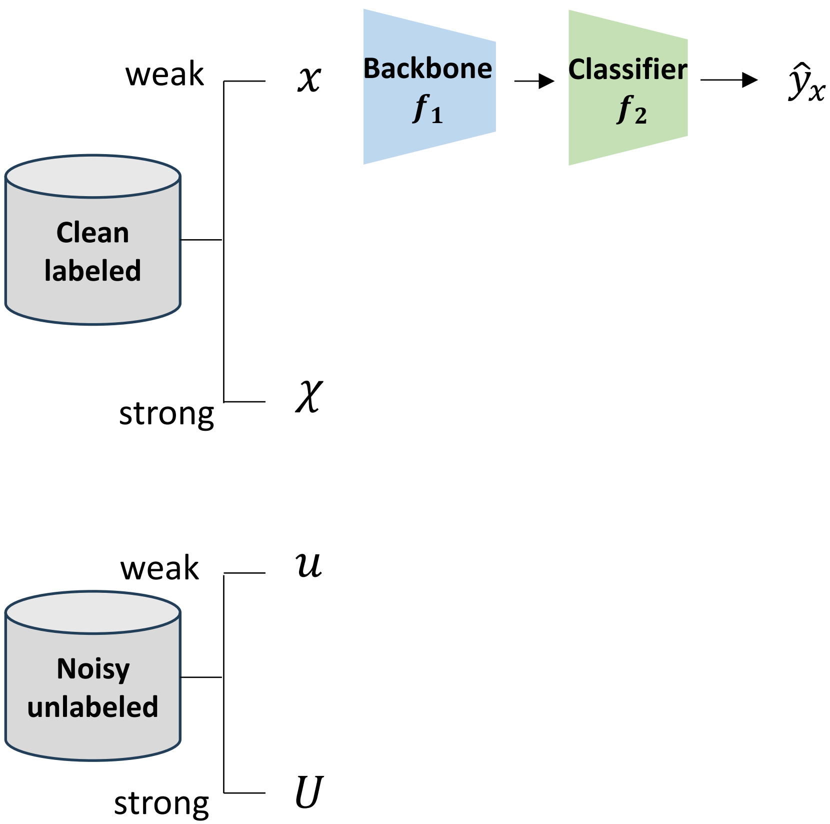
Manifold DivideMix



Method

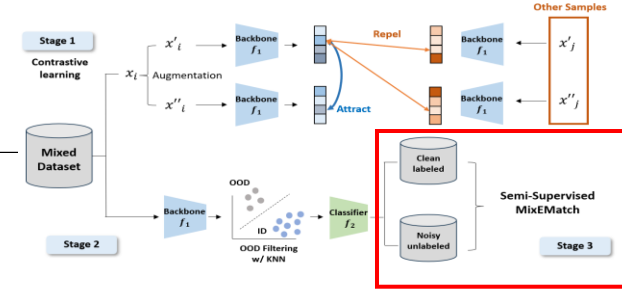
❖ Semi-Supervised Learning with MixEMatch

- Weak augmented 데이터를 통해 label refinement / pseudo-labeling



DMQA

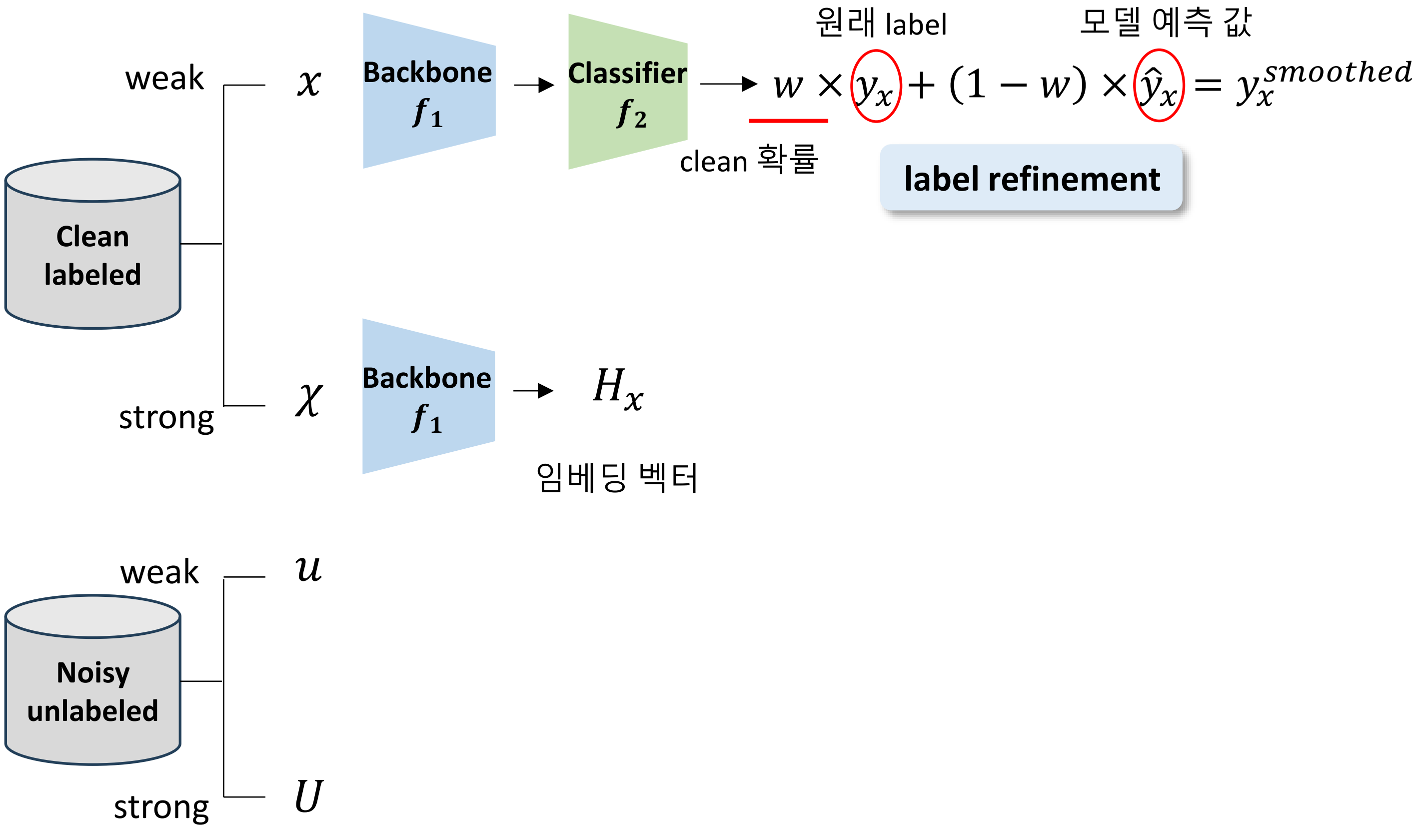
Manifold DivideMix



Method

❖ Semi-Supervised Learning with MixEMatch

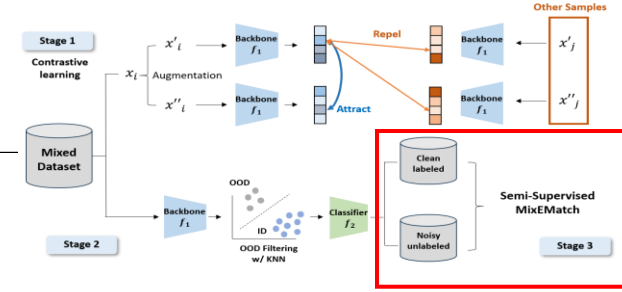
- Weak augmented 데이터를 통해 label refinement / pseudo-labeling



DMQA

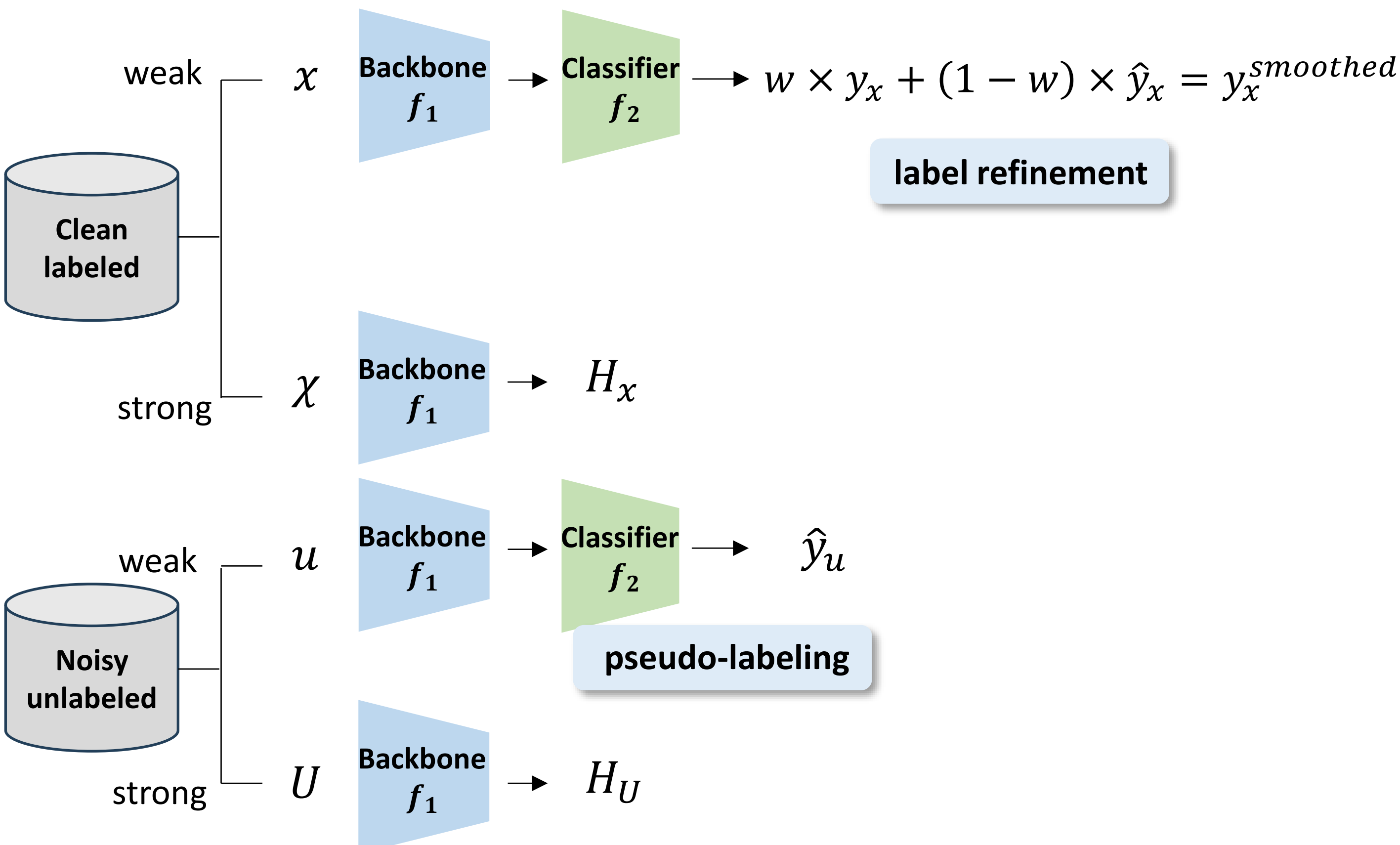
Manifold DivideMix

Method



❖ Semi-Supervised Learning with MixEMatch

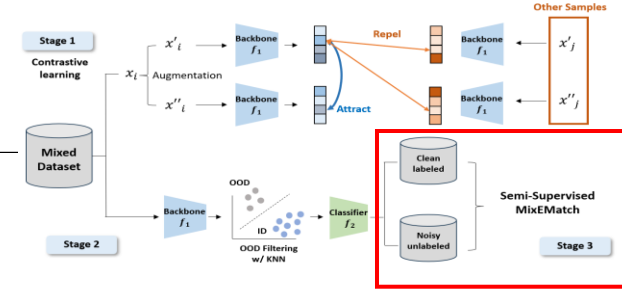
- Weak augmented 데이터를 통해 label refinement / pseudo-labeling



DMQA

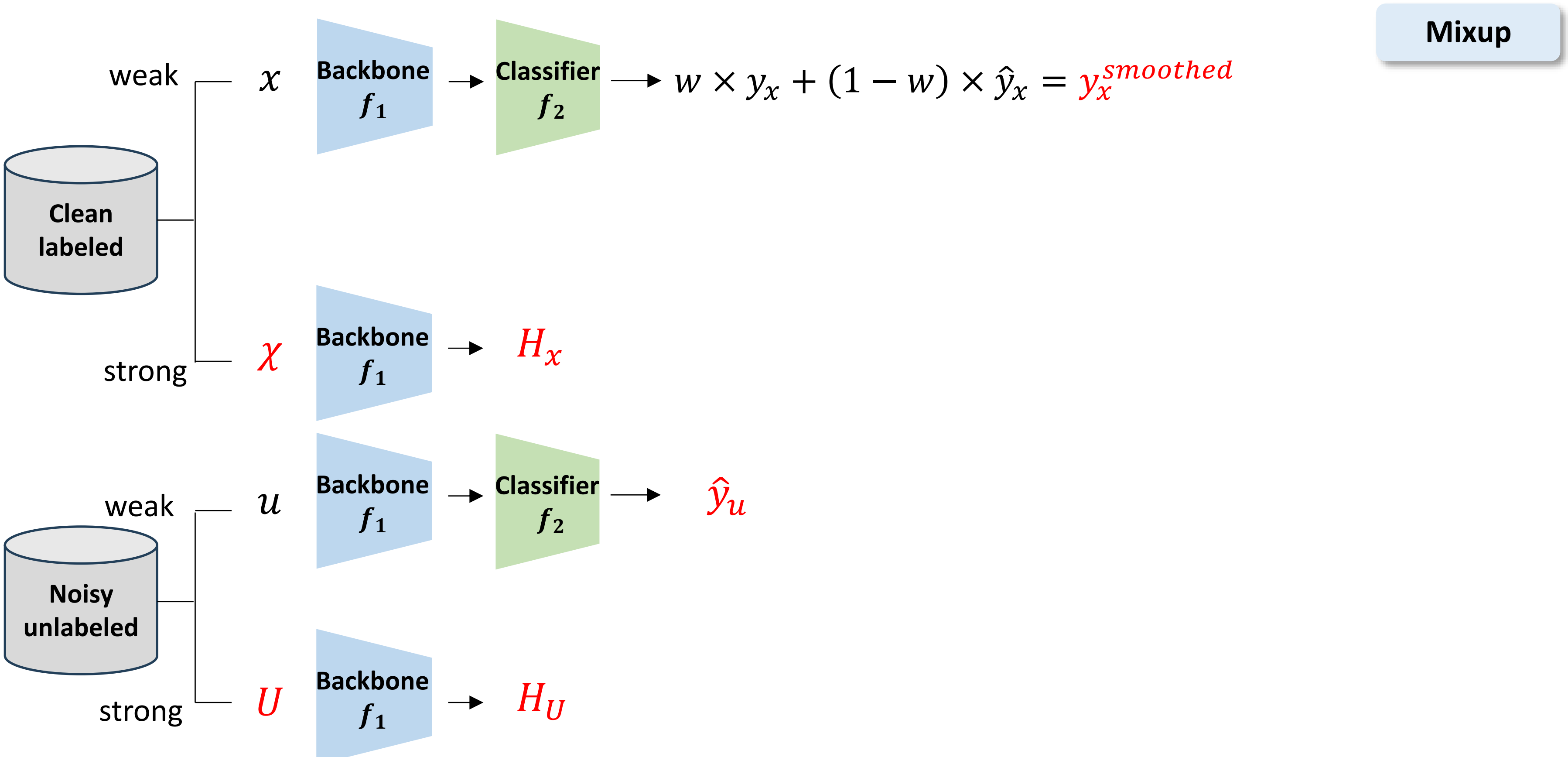
Manifold DivideMix

Method



❖ Semi-Supervised Learning with MixEMatch

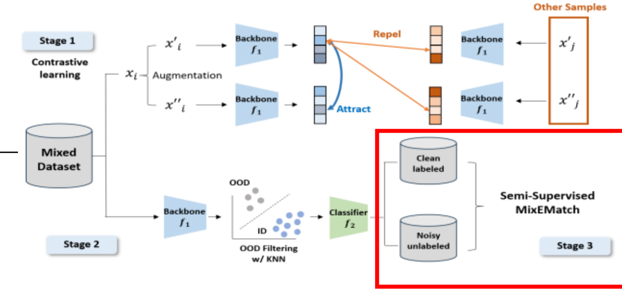
- 인풋, 아웃풋, 임베딩 벡터 값에 대해 MixMatch 적용



DMQA

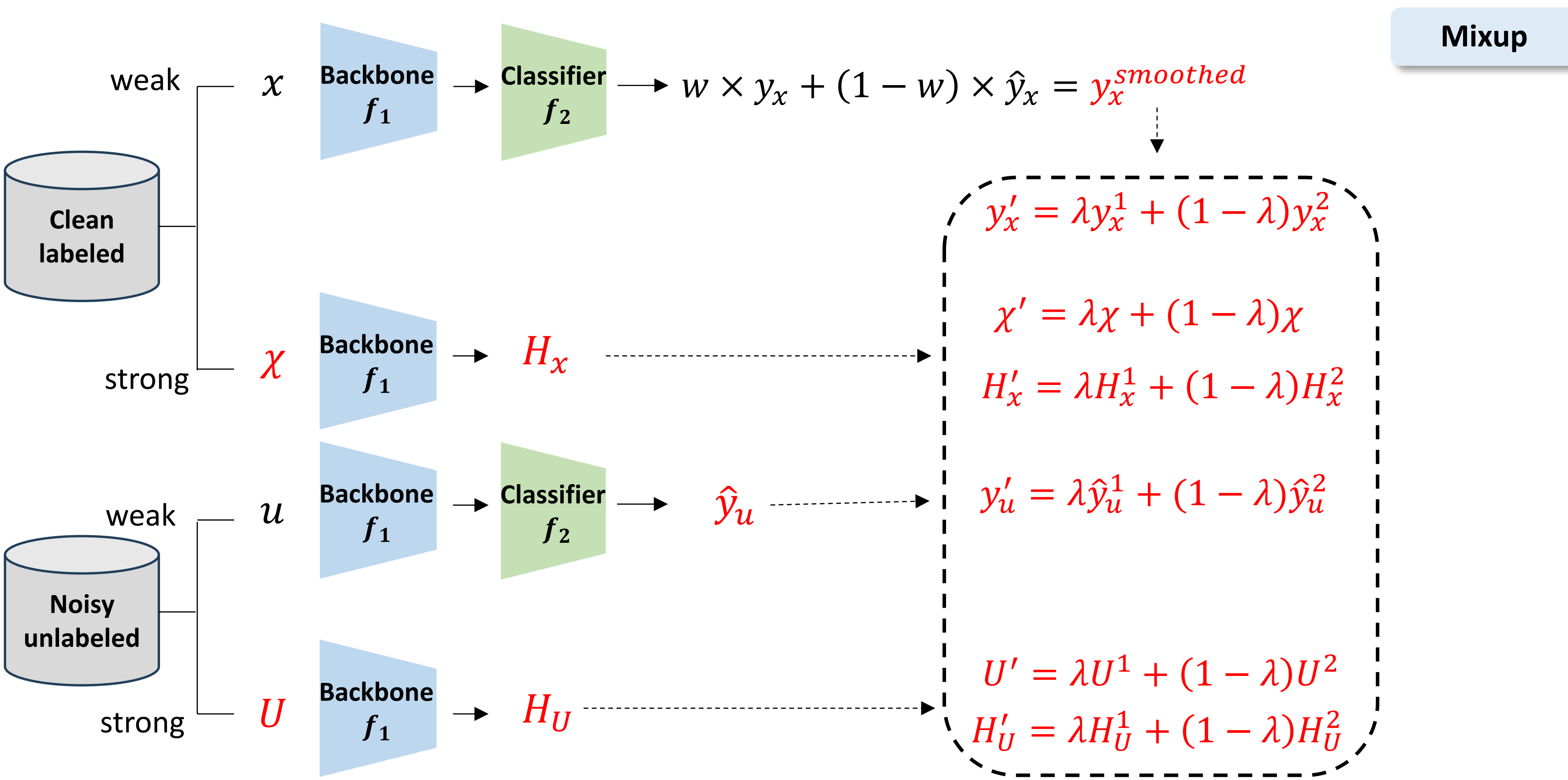
Manifold DivideMix

Method



❖ Semi-Supervised Learning with MixEMatch

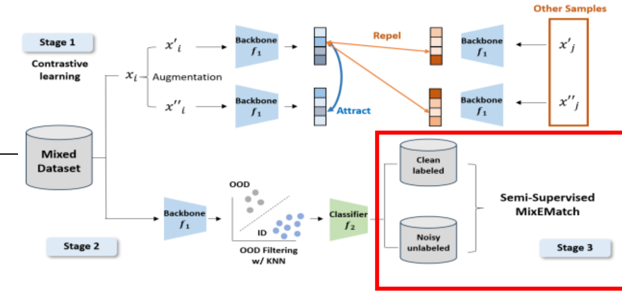
- 인풋, 아웃풋, 임베딩 벡터 값에 대해 MixMatch 적용



DMQA

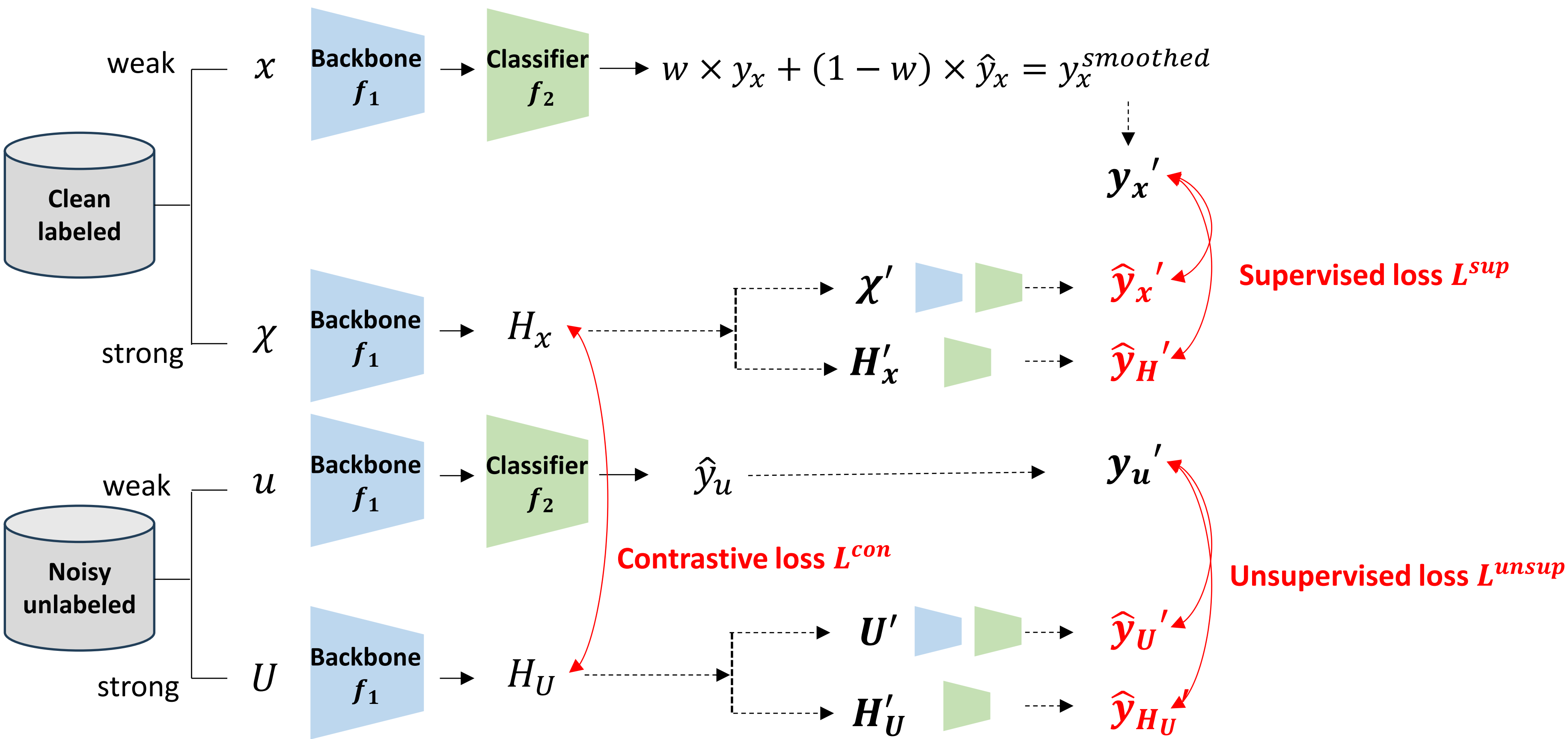
Manifold DivideMix

Method



❖ Semi-Supervised Learning with MixEMatch

- $L = L^{sup} + \lambda_u \times L^{unsup} + \lambda_c \times L^{con}$



DMQA

Manifold DivideMix

Experiments

❖ Main results

- ID 데이터셋: CIFAR100 / OOD 데이터셋: ImageNet32, Places365
- 모든 noise 세팅에서 가장 우수한 성능 도출

Corruption	ImageNet32				Places365			
Noise Ratio (r_{out},r_{in})	20%, 20%	40%, 20%	60%, 20%	40%, 40%	20%, 20%	40%, 20%	60%, 20%	40%, 40%
CE	55.5	44.3	26.0	18.5	53.6	42.5	21.4	13.9
Mixup [41]	62.5	53.2	40.4	33.9	59.7	48.6	33.7	27.6
JoSRC [38]	64.2	61.4	37.1	41.4	66.7	60.6	39.6	32.6
ELR [23]	68.5	63.0	44.6	34.2	68.5	62.3	36.5	33.9
EDM [27]	70.4	61.8	14.6	1.6	70.3	61.6	14.7	11.9
DSOS [2]	70.5	62.1	49.1	42.9	69.1	59.5	35.4	29.5
RRL [20]	72.3	65.4	24.5	30.6	72.5	65.8	49.3	24.3
SNCP [1]	72.7	67.1	51.3	52.7	71.1	63.5	49.8	47.6
Ours	**75.3**	**69.1**	**59.8**	**63.0**	**74.5**	**69.3**	**59.0**	**62.5**

Table 1. Comparison of classification accuracy with the state-of-the-art methods on CIFAR100 corrupted with ImageNet32 or Places365 images with ID and OOD noise of ratio of r_{in} and r_{out}, respectively. Results of other previous methods are from [1].

DMQA

Manifold DivideMix

Experiments

❖ Real-world noisy dataset results

1. ID 데이터셋: miniImageNet / OOD 데이터셋: CNWL
   - ➢ miniImageNet: Real-world noisy dataset
2. Clothing1M / WebVision: Real-world noisy dataset

Method	Noise Ratio 20%	40%	60%	80%
CE	47.4	42.7	37.3	29.8
Mixup [41]	49.1	46.4	40.6	33.6
DivideMix [19]	51.0	46.7	43.1	34.5
ScanMix [28]	59.1	54.5	52.4	40.0
PropMix [8]	61.2	56.2	52.8	43.4
SNCP [1]	61.6	59.9	54.9	45.6
PLS [3]	63.1	60.0	54.4	46.5
Ours	**64.4**	**61.4**	**56.2**	**47.8**

Table 2. Comparison of classification accuracy with the state-of-the-art methods on Web-corrupted miniImageNet from the CNWL [15] (32 × 32).

Method	Clothing1M	(mini)WebVision	
	Top1	Top1	Top5
GCE [43]	71.7	61.2	80.8
SL [34]	72.1	63.8	84.3
ELR+ [23]	71.5	63.6	83.5
Co-teaching [11]	72.5	64.1	85.0
JoCor [38]	71.7	60.8	82.5
DivideMix [19]	74.6	77.2	91.6
GSS-SSL [40]	**74.9**	77.4	**93.1**
Ours	73.1	**78.4**	92.0

Table 3. Comparison of classification accuracy with the state-of-the-art methods on Clothing1M and (mini)Webvision.

Manifold DivideMix

Experiments

❖ Random noise results

- CIFAR10 dataset에 symmetric / asymmetric noise 주입 후 실험

	Symmetric Noise				Asymmetric Noise		
Method	20%	50%	80%	90%	10%	30%	40%
CE	86.8	79.4	62.9	42.7	88.8	81.7	76.1
MixUp	95.6	87.1	71.6	52.2	93.3	83.3	77.7
DivideMix	95.0	93.7	92.4	74.2	93.8	92.5	91.4
ELR	95.8	94.8	93.3	78.7	95.5	94.8	93.0
UNICON	96.0	95.6	93.9	88.1	95.3	94.8	94.1
Propmix	**96.1**	95.5	93.7	93.2	-	-	**94.6**
GSS-SSL	94.3	-	91.6	-	-	92.4	91.8
Ours	96.0	**95.7**	**94.6**	**93.7**	**95.6**	**95.0**	93.8

Table 4. Comparison of classification accuracy on CIFAR10 with ID symmetric and asymmetric noise.

Manifold DivideMix

Experiments

❖ Ablation study

- ID 데이터셋: CIFAR100 / OOD 데이터셋: ImageNet32

		FixMatch	MixEMatch	Best	Last
No noise corr	CE	✗	✗	41.4	18.5
	SSL+LC	✗	✗	47.5	19.9
Noise Robust	ID	✓	✗	58.6	58.3
	ID + OOD	✓	✗	58.0	58.0
	ID	✗	✓	61.6	61.4
	ID + OOD	✗	✓	**63.1**	**63.0**

Table 5. Ablation study on CIFAR-100 corrupted with ImageNet32 with $r_{out} = 40\%$ and $r_{in} = 40\%$.

Manifold DivideMix

Experiments

❖ ID / OOD noise detection

- ID 데이터셋: CIFAR100 / OOD 데이터셋: ImageNet32

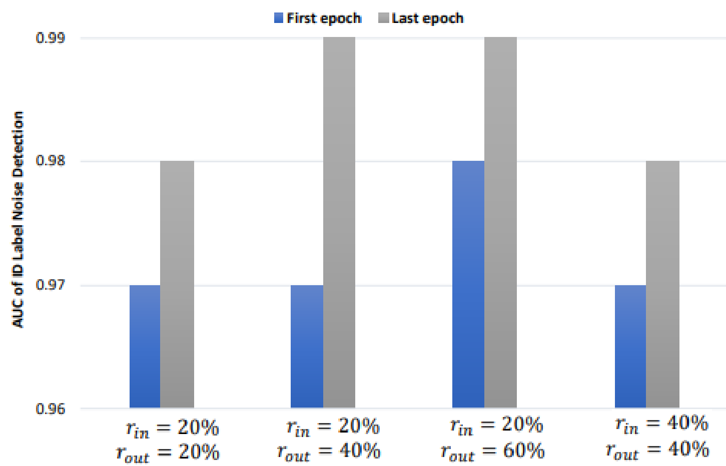


Figure 1. AUC of clean/ID labeled noise detection using GMM on training data of CIFAR-100 corrupted with ImageNet32. First epoch means the first epoch of our Semi-Supervised learning step (first epoch after warmup training) and last epoch means the final last epoch of semi-supervised learning.

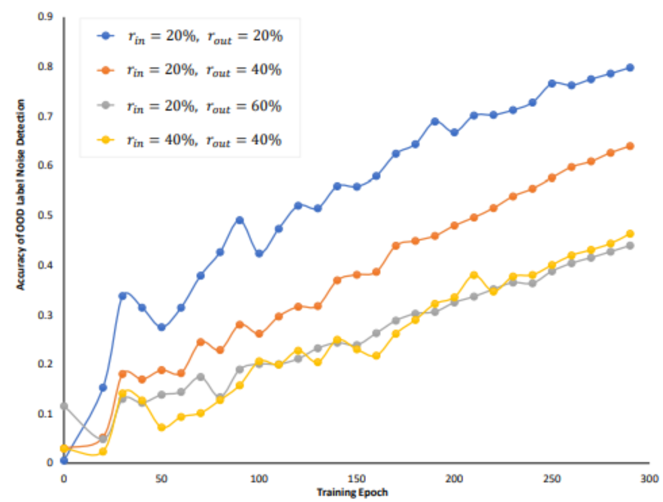


Figure 2. Accuracy of OOD labeled noise detection during warmup and semi-supervised learning steps for different level of in- and out-of-distribution labeled noise.

[ID noise detection]

[OOD noise detection]

DMQA

Conclusion

Conclusion

❖ **DivideMix**

- Label noise learning에 MixMatch를 최초로 적용하였으며 후속 연구들의 뿌리가 된 연구

❖ **Bayesian DivideMix++**

- DivideMix에 불확실성을 적용하여 보다 강건하게 학습하도록 함

❖ **Manifold DivideMix**

- OOD noise까지 결합된 문제 상황
- Embedding 벡터까지 mixup하는 MixEMatch를 제안하였으며, 잠재적인 ood noise를 잘 구분할 수 있도록 함

DMQA

References

DMQA

References

- Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., & Raffel, C. A. (2019). Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, *32*.

- Li, J., Socher, R., & Hoi, S. C. (2020). Dividemix: Learning with noisy labels as semi-supervised learning. arXiv preprint arXiv:2002.07394. Izacard, G., Caron, M., Hosseini, L., Riedel, S., Bojanowski, P., Joulin, A., & Grave, E. (2021).

- Nagarajan, B., Marques, R., Aguilar, E., & Radeva, P. (2024). Bayesian DivideMix++ for enhanced learning with noisy labels. Neural Networks, 172, 106122.

- Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (1805). Autoaugment: Learning augmentation policies from data. arXiv 2018. *arXiv preprint arXiv:1805.09501*, *2*.

- Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 702-703).

- Fooladgar, F., To, M. N. N., Mousavi, P., & Abolmaesumi, P. (2024). Manifold DivideMix: A semi-supervised contrastive learning framework for severe label noise. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4012-4021).

DMQA

고맙습니다