Contents

A. List of Symbols

This section summarizes the list of symbols used in this paper:

Notations	Explanation
$\mathbb{R}^d_{>1}$	d-dimensional real value space with all values larger than 1
$\mathcal{X} \subseteq \mathbb{R}^d$	Data space as subsset of the d-dimensional real value set
$\mathcal{Y} = \{1, 2, ..., K\}$	Label space of K ID classes
$\mathcal{Y} \cup \{K+1\}$	Label space of K ID classes and one $K+1$ OOD class
X_s, Y_s, B_s	Random Variable of the source domain image (X_s), label (Y_s) and ID data indicator (B_s)
X_t, Y_t, B_t	Random Variable of the target domain image (X_t), label (Y_t) and ID data indicator (B_t)
$p_s(\cdot)$	Source domain distribution (*e.g.* $p_s(x)$ for X_s, $p_s(b = \cdot)$ for B_s)
$p_t(\cdot)$	Tource domain distribution (*e.g.* $p_t(x)$ for X_t, $p_t(y = \cdot)$ for Y_t)
$\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$	Source domain labelled dataset with N_s samples
$\mathcal{D}^t = \{x_i^t\}_{i=1}^{N_t}$	Target domain unlabelled dataset with N_t samples
$\mathcal{D}^{\mathbf{o}} = \{x_i^o\}_{i=1}^{N_o}$	OOD reference dataset with N_o samples
$\mathbf{c}, \rho_s$	Source domain ID label distribution $p_s(y = \cdot) = \mathbf{c}$ and ID data ratio $p_s(b = 1) = \rho_s$
$\boldsymbol{\pi}, \rho_t$	Target domain ID label distribution $p_t(y = \cdot) = \boldsymbol{\pi}$ and ID data ratio $p_t(b = 1) = \rho_t$
$f : \mathcal{X} \rightarrow \Delta^{K-1}$	Source domain ID classifier that output K dimensional probability simplex
$h : \mathcal{X} \rightarrow [0, 1]$	Source domain ID vs OOD classifier that output a scalar in $[0, 1]$
γ, T	Hyperparameters of our model when use Gaussian noise to generate pseudo OOD samples
$\lvert\mathcal{D}\rvert$	Returns cardinality of the dataset $\mathcal{D}$, *i.e.* $\lvert\mathcal{D}\rvert = N$ if $\mathcal{D} = \{x_i\}_{i=1}^{N}$

B. Related Works

B.1. MLLS

The Maximum Likelihood Label Shift method is a closed set label shift estimation model that was originally proposed by Saerens et al. 47. With unlabeled target domain data $\mathcal{D}^t = \{x_i^t\}_{i=1}^{N^t}$, MLLS estimates the target label distribution $p_t(y = \cdot) = \boldsymbol{\pi}$ by maximizing the log likelihood:

$$\log L(\boldsymbol{\pi}; \mathcal{D}^t) := \log \left(\prod_{i=1}^{N^t} \sum_{j=1}^{K} \frac{\pi_j}{c_j} f(x_i)_j \right)$$

using the EM algorithm, which is stated in Alg. 3.

Algorithm 3 MLLS

Input: $\mathcal{D}^t = \{x_i^t\}_{i=1}^{N^t}, p_s(y = \cdot) = \mathbf{c}, f : \mathcal{X} \to \Delta^{K-1}$.
Initialize: $\boldsymbol{\pi}^{(0)} \in \Delta^{K-1}$.
for $m = 0$ to M **do**
  E-step: Evaluate

$$g_{ij}^m = \frac{\frac{\pi_j}{c_j} f(x_i^t)_j}{\sum_{l=1}^{K} \frac{\pi_l}{c_l} f(x_i^t)_l}.$$

  M-step: Evaluate

$$\pi_j^{(m+1)} = \frac{1}{N^t} \sum_{i=1}^{N^t} g_{ij}^m.$$

end for
Output: $p_t(y = \cdot) = \boldsymbol{\pi}^{(M+1)}$.

The iterative procedure of the EM algorithm is repeated until numerical convergence to obtain the MLE of the target label distribution $\boldsymbol{\pi}^{\text{MLE}}$, which satisfies:

$$\boldsymbol{\pi}^{\text{MLE}} \in \underset{\boldsymbol{\pi} \in \Delta^{K-1}}{\arg\min} - \log L(\boldsymbol{\pi}; \mathcal{D}^t).$$

In the following works, Alexandari et al. 1 proves that the NLL objective of MLLS is convex and empirically demonstrates that MLLS outperform other closed set label shift estimation methods in many image classification datasets. Garg et al. 12 proved that MLLS is consistent when classifier f is canonically calibrated.

B.2. MAPLS

The Maximum *a Posteriori* Label Shift (MAPLS) method is also a closed set label shift estimation model, which was recently proposed by Ye et al. 62. By introducing a Dirichlet prior $\boldsymbol{\pi} \sim \text{Dir}(K, \boldsymbol{\alpha})$ over the target label distribution $\boldsymbol{\pi}$, MAPLS aims at optimizing the posterior:

$$p(\boldsymbol{\pi} | \mathcal{D}^t, \boldsymbol{\alpha}) = \frac{1}{Z} \prod_{i=1}^{K} \pi_i^{\alpha_i - 1} \prod_{i=1}^{N^t} \sum_{j=1}^{K} \frac{\pi_j}{c_j} f(x_i)_j,$$

where Z is the normalization constant. The EM algorithm of MAPLS can be written as Alg. 4.

Ye et al. 62 further proved that the optimization objective of MAPLS algorithm is strictly convex and EM algorithm is guaranteed to converge to the MAP estimate $\boldsymbol{\pi}^{\text{MAP}}$ which satisfies:

$$\boldsymbol{\pi}^{\text{MAP}} = \underset{\boldsymbol{\pi} \in \Delta^{K-1}}{\arg\min} - \frac{1}{Z} \prod_{i=1}^{K} \pi_i^{\alpha_i - 1} \prod_{i=1}^{N^t} \sum_{j=1}^{K} \frac{\pi_j}{c_j} f(x_i)_j.$$

The author of the MAPLS algorithm also empirically demonstrates that MAPLS outperforms other closed set label shift estimation models in large scale datasets like ImageNet, especially under large label shift settings.

Algorithm 4 MAPLS

Input: $\mathcal{D}^t = \{x_i^t\}_{i=1}^{N^t}$, $p_s(y=\cdot) = \mathbf{c}$, $f : \mathcal{X} \to \Delta^{K-1}$, $\boldsymbol{\alpha} \in \mathbb{R}_{>1}^K$.
Initialize: $\boldsymbol{\pi}^{(0)} \in \Delta_{>0}^{K-1}$.
for $m = 0$ to M **do**
  E-step Evaluate:

$$g_{ij}^m = \frac{\frac{\pi_j^{(m)}}{c_j} f(x_i)_j}{\sum_{l=1}^K \frac{\pi_l^{(m)}}{c_l} f(x_i)_l}. \tag{18}$$

  M-step Obtain $\boldsymbol{\pi}^{(m+1)}$ with:

$$\pi_j^{(m+1)} = \frac{\sum_{i=1}^{N^t} g_{ij}^m + \alpha_j - 1}{N^t + \sum_{l=1}^K (\alpha_l - 1)}. \tag{19}$$

end for
Output: $p_t(y=\cdot) = \boldsymbol{\pi}^{(M+1)}$.

B.3. OOD detection

OOD detection has been widely studied in the Deep Learning regime. Existing approaches can be categorized into roughly three categories: post-hoc inference methods and training methods with or without OOD data.

The majority of the OOD methods are post-hoc inference methods, where the OOD classifier is constructed based on a pre-trained classifier over ID classes. OpenMax [4] proposed to construct the OOD classifier by modelling per-class features with a Weibull distribution. MSP [21] utilize the maximal softmax score of the ID classifier prediction. ODIN [31] observed that NN models respond to ID and OOD data differently under adversarial attacks [16]. MDS [30] also adopts the adversarial attack approach but detects OOD data with a Mahalanobis distance-based score. OpenGAN [27] trains an extra discriminator network to distinguish ID and OOD features. EBO [36] proposed an Energy-based score to detect OOD samples. GRAM [49] establish their model with Gram matrices. ReAct [50] demonstrates that rectifying the penultimate layer features of the pre-trained classifier can help post-hoc OOD detection methods. MLS [22] argues that maximal logit score is a better OOD indicator. VIM [57] propose a three stage pipline to compute the OOD score by adjusting the features, logits and softmax probability of the ID classifier. Sun et al. 51 introduces a k-Nearest Neighbor (KNN) based OOD classifier. Ash [8] shows that pruning image features in the intermediate layers can help OOD detection.

Among training methods without OOD data, Hendrycks et al. 24 argues that training the classifier with an auxiliary self-supervised rotation loss is beneficial to OOD detection models. GODIN [25] extends the ODIN model by introducing an extra linear layer that models the probability the data is not OOD given the image. CSI [52] enhances a baseline OOD detector by training the classifier with a loss that contrasts ground truth samples with distribution shifted samples. APRL [6] encourages ID samples to move far away from a bounded space left for OOD data.

In the machine learning community, Miller et al. 40, Vaze et al. 54 argue that a good ID classifier implies a good OOD classifier. Hein et al. 20 shows that for OOD sample, a ReLU network can predict its label as ID class with arbitrary high confidence. Meinke and Hein 38 propose a GMM based classifier approach to prevent the model from assigning OOD data with high confidence. Fang et al. 11 analyzes the conditions under which OOD detection is learnable.

B.4. Open Set Domain Adaptation

Compared with the Open Set Label Shift problem, the Open Set Domain Adaptation (OSDA) task considers a slightly different setup. Theoretically, the OSDA task does not require the Label Shift Assumption 3.2 to hold between the source and target domain. Empirically, OSDA models focus more on tackling the image distribution $p(x)$ shift rather than the label distribution $p(y)$ shift, where they are usually tested with source and target domain having identical ID label distribution [43].

Similar to the relation between Closed Set Label Shift and Open Set Label Shift, the OSDA task extends the Closed Set Domain Adaptation (CSDA) task by allowing target domain having an extra class that contains all the new categories that not appear in the source domain [10, 33, 43, 48, 58, 65]. As the OSDA problem setup and the OSDA models are not used in this paper, we will not go into details of the OSDA problem. Therefore, we would like to refer readers who are interested in the OSDA problem to the cited literature for detailed discussions.

C. Mathematical Proofs

C.1. Proof of Theorem 4.1 (See page 4)

Theorem 4.1. *(**Source ID/OOD ratio estimator**) Under Assumption 3.3B, given source ID dataset $\mathcal{D}^s$ and source OOD dataset $\mathcal{D}^{\mathbf{o}}$, then for $\delta > 0$, with probability of at least $1-2\delta$,*

$$|\rho_s - \hat{\rho}_s| \leq \frac{1}{1-\mu_1+\mu_0}\sqrt{\frac{\log 1/\delta}{2\min(|\mathcal{D}^{\mathbf{o}}|, |\mathcal{D}^s|)}} \tag{4}$$

where $\mu_0 := \mathbb{E}_{X_s|B_s=0}[h(x)]$, $\mu_1 := \mathbb{E}_{X_s|B_s=1}[h(x)]$.

Proof. Given the available information, for $p_s(b=1) = \rho_s$ we have:

$$\begin{aligned} p_s(b=1) &= \mathbb{E}_{X_s}[p(b=1|x)] = \mathbb{E}_{X_s}[h(x)] = \mathbb{E}_{B_s}[\mathbb{E}_{X_s|B_s}[h(x)]] \\ &= (1-p_s(b=1)) \cdot \mathbb{E}_{X_s|B_s=0}[h(x)] + p_s(b=1) \cdot \mathbb{E}_{X_s|B_s=1}[h(x)] \end{aligned} \tag{20}$$

Rearranging the equation and we can get:

$$\begin{aligned} \rho_s &= \frac{\mathbb{E}_{X_s|B_s=0}[h(x)]}{1-\mathbb{E}_{X_s|B_s=1}[h(x)] + \mathbb{E}_{X_s|B_s=0}[h(x)]} \\ &= \frac{\mu_0}{1-\mu_1+\mu_0}, \end{aligned} \tag{21}$$

where $\mu_0 := \mathbb{E}_{X_s|B_s=0}[h(x)]$ and $\mu_1 := \mathbb{E}_{X_s|B_s=1}[h(x)]$.

The expectation terms can be approximated given OOD dataset $\mathcal{D}^{\mathbf{o}}$ and source ID dataset $\mathcal{D}^s$:

$$\begin{cases} \mathbb{E}_{X_s|B_s=0}[h(x)] \approx \frac{1}{|\mathcal{D}^{\mathbf{o}}|}\sum_{x\in\mathcal{D}^{\mathbf{o}}} h(x) \\ \mathbb{E}_{X_s|B_s=1}[h(x)] \approx \frac{1}{|\mathcal{D}^{s}|}\sum_{x\in\mathcal{D}^{s}} h(x), \end{cases} \tag{22}$$

which yields the approximation $\hat{\rho}$:

$$\hat{\rho} = \frac{\hat{\mu}_0}{1-\hat{\mu}_1+\hat{\mu}_0}, \tag{23}$$

where $\hat{\mu}_0 := \frac{1}{|\mathcal{D}^{\mathbf{o}}|}\sum_{x\in\mathcal{D}^{\mathbf{o}}} h(x)$ and $\hat{\mu}_1 := \frac{1}{|\mathcal{D}^{s}|}\sum_{x\in\mathcal{D}^{s}} h(x)$.

Note that since $h(x) \in [0,1]$, Hoeffding's inequality [55] guarantees for all $\epsilon > 0$:

$$\begin{aligned} p\left(|\mu_0 - \hat{\mu}_0| \geq \epsilon\right) \leq 2e^{-2|\mathcal{D}^{\mathbf{o}}|\epsilon^2} \\ p\left(|\mu_1 - \hat{\mu}_1| \geq \epsilon\right) \leq 2e^{-2|\mathcal{D}^{s}|\epsilon^2}. \end{aligned} \tag{24}$$

Therefore with high probability of at least $1 - 2e^{-2\min(|\mathcal{D}^{\mathbf{o}}|,|\mathcal{D}^s|)\epsilon^2}$ we have:

$$\begin{cases} \rho - \hat{\rho} \leq \frac{\mu_0}{1-\mu_1+\mu_0} - \frac{\mu_0+\epsilon}{1-(\mu_1+\epsilon)+(\mu_0+\epsilon)} = \frac{\epsilon}{1-\mu_1+\mu_0}, \\ \rho - \hat{\rho} \geq \frac{\mu_0}{1-\mu_1+\mu_0} - \frac{\mu_0-\epsilon}{1-(\mu_1-\epsilon)+(\mu_0-\epsilon)} = \frac{-\epsilon}{1-\mu_1+\mu_0}, \end{cases} \tag{25}$$

for all $\delta \in [0, \max((1-\mu_0)/2, (1-\mu_1)/2)]$, which is equivalent to:

$$|\rho - \hat{\rho}| < \frac{\epsilon}{1-\mu_1+\mu_0}. \tag{26}$$

Letting $\delta := e^{-2\min(|\mathcal{D}^{\mathbf{o}}|,|\mathcal{D}^s|)\epsilon^2}$, rearrange the equations and we get the result. □

C.2. Extension of Theorem 4.1 to the Multi-Class setting (See page 4)

Problem Setup: (General) Given a blackbox model $h : \mathcal{X} \to \Delta^{K-1}$ that satisfies $h(x) = p(y|x)$ for distribution $p(x, y)$ and K datasets $\mathcal{D}^1, \mathcal{D}^2, ..., \mathcal{D}^K$, with $\mathcal{D}^k$ containing samples drawn i.i.d. from $p(x|y = k)$, we want to estimate the label distribution $p(y = \cdot) = \boldsymbol{\rho} \in \Delta^{K-1}$.

Similar to the binary case, we can write the label distribution as the sum of the conditional expectation:

$$\begin{aligned}\rho_j &= \mathbb{E}_X[p(y = j|x)] = \mathbb{E}_X[h(x)_j] = \mathbb{E}_Y[\mathbb{E}_{X|Y}[h(x)_j]] \\ &= \sum_{k=1}^{K} p(y = k)\mathbb{E}_{X|Y}[h(x)_j] = (\mu\boldsymbol{\rho})_j,\end{aligned} \tag{27}$$

where $\mu \in \mathbb{R}^{K\times K}$ with $\mu_{jk} := \mathbb{E}_{X|Y=k}[h(x)_j]$.

Lemma C.1. *(**Multi-Class**) If $p(y|x) = h(x)$, given $\mathcal{D}^1, \mathcal{D}^2, ..., \mathcal{D}^K$ containing samples x drawn i.i.d. from $p(x|y = 1), p(x|y = 2), ..., p(x|y = K)$, then for $p(y) = \boldsymbol{\rho} \in \Delta^{K-1}$ we have:*

$$\underset{\boldsymbol{\rho}\in\Delta^{K-1}}{\arg\min} \|(\hat{\mu} - \mathbf{I})\boldsymbol{\rho}\|_2^2 \xrightarrow[a.s.]{} \boldsymbol{\rho}, \tag{28}$$

where $\hat{\mu} \in \mathbb{R}^{K\times K}$ is a stochastic matrix with $\mu_{jk} = \frac{1}{|\mathcal{D}^k|} \sum_{x\in\mathcal{D}^k} h(x)_j$.

Proof. Given the available information, let $p(y = j) = \rho_j$ for all $j \in \mathcal{Y} = \{1, 2, ..., K\}$, then we have:

$$\begin{aligned}\rho_j &= \mathbb{E}_X[p(y = j|x)] = \mathbb{E}_X[h(x)_j] = \mathbb{E}_Y[\mathbb{E}_{X|Y}[h(x)_j]] \\ &= \sum_{k=1}^{K} p(y = k)\mathbb{E}_{X|Y}[h(x)_j] \\ &= (\mu\boldsymbol{\rho})_j,\end{aligned} \tag{29}$$

where $\mu \in \mathbb{R}^{K\times K}$ with $\mu_{jk} := \mathbb{E}_{X|Y=k}[h(x)_j]$.

The μ can be approximated via:

$$\mu_{jk} \approx \hat{\mu}_{jk} := \frac{1}{|\mathcal{D}^k|} \sum_{x\in\mathcal{D}^k} h(x)_j. \tag{30}$$

Hence we can approximate $\boldsymbol{\rho}$ with $\hat{\boldsymbol{\rho}}$ that is defined as:

$$\hat{\boldsymbol{\rho}} := \underset{\boldsymbol{\rho}\in\Delta^{K-1}}{\arg\min} \|(\hat{\mu} - \mathbf{I})\boldsymbol{\rho}\|_2^2. \tag{31}$$

□

C.3. Proof of Lemma 4.2 (See page 5)

Lemma 4.2. *Under Assumption 3.2, 3.3, given $\mathcal{D}^t$, the negative log likelihood $-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t)$ can be written as:*

$$-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) = -\sum_{i=1}^{N^t} \log\left(\sum_{j=1}^{K+1} \frac{\tilde{\pi}_j}{\tilde{c}_j} \tilde{f}(x_i)_j\right) + C, \tag{5}$$

where C does not depend on either $\boldsymbol{\pi}$ or ρ_t and

$$\tilde{f}(x)_i := \begin{cases} h(x) \cdot f(x)_i, & i \in \mathcal{Y} \\ 1 - h(x), & i = K + 1, \end{cases}, \tag{6}$$

$$\begin{aligned}\tilde{\boldsymbol{\pi}} &:= [\rho_t \cdot \pi_1, ..., \rho_t \cdot \pi_K, 1 - \rho_t]^T \\ \tilde{\mathbf{c}} &:= [\rho_s \cdot c_1, ..., \rho_s \cdot c_K, 1 - \rho_s]^T.\end{aligned} \tag{7}$$

Proof. The label shift assumption can be written as:

$$p_s(x|y=i) = p_t(x|y=i) \quad \text{for all} \quad i \in \mathcal{Y} \cup \{K+1\} \tag{32}$$

On target domain, if we are given only unlabeled images $\mathcal{D}^t = \{x_i^t\}_{i=1}^{N^t}$, we can construct the likelihood:

$$L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) = \prod_{i=1}^{N^t} p_t(x; \boldsymbol{\pi}, \rho_t) = \prod_{i=1}^{N^t} \left(\sum_{l=1}^{2} \sum_{j=1}^{K} p_t(x_i|y=j) p_t(y=j|b=l) p_t(b=l) \right). \tag{33}$$

Note that $p_s(b=1) = \rho_s$, $p_t(b=1) = \rho_t$ and in Eq. 1 for all $(x, j) \in \mathcal{X} \times (\mathcal{Y} \cup \{K+1\})$ we have:

$$p_s(y|b; \mathbf{c}) = \begin{cases} c_j, & \text{if } b=1, y \neq K+1 \\ 1, & \text{if } b=0, y=K+1 \\ 0, & \text{otherwise} \end{cases}, \qquad p_s(y|b; \mathbf{c}) = \begin{cases} \pi_j, & \text{if } b=1, y \neq K+1 \\ 1, & \text{if } b=0, y=K+1 \\ 0, & \text{otherwise} \end{cases}. \tag{34}$$

Based on Eq. (34) and label shift Assumption 3.2 we have:

$$\begin{aligned} L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) &= \prod_{i=1}^{N^t} \left(\sum_{l=1}^{2} \sum_{j=1}^{K} p_t(x_i|y=j) p_t(y=j|b=l) p_t(b=l) \right) \\ &= \prod_{i=1}^{N^t} \left(\sum_{j=1}^{K} p_t(x_i|y=j) p_t(y=j|b=1) p_t(b=1) + p_t(x_i|y=K+1) p_t(y=K+1|b=0) p_t(b=0) \right) \\ &= \prod_{i=1}^{N^t} \left(\sum_{j=1}^{K} p_s(x_i|y=j) p_t(y=j|b=1) p_t(b=1) + p_s(x_i|y=K+1) p_t(y=K+1|b=0) p_t(b=0) \right) \\ &= \prod_{i=1}^{N^t} \left(\sum_{j=1}^{K} \frac{p_s(y=j|x_i)}{p_s(y=j)} p_t(y=j|b=1) p_t(b=1) + \frac{p_s(y=K+1|x_i)}{p_s(y=K+1)} p_t(y=K+1|b=0) p_t(b=0) \right) \\ &\quad \cdot Const, \end{aligned} \tag{35}$$

where $Const := \prod_{i=1}^{N^t} p_s(x_i)$ is irrelevant to $\boldsymbol{\pi}$ or ρ_t.

Based on Eq. (34) and Assumption 3.3: $p_s(y=\cdot|x, b=1) = f(x)$ and $p_s(b=1|x) = h(x)$ we have:

$$\begin{aligned} 1 - h(x) &= p_s(b=0|x_i) \\ &= \sum_{i=1}^{K+1} p_s(b=0|y=i) p(y=i|x_i) \\ &= \sum_{i=1}^{K+1} \frac{p_s(y=i|b=0) p_s(b=0)}{p_s(y=i)} p(y=i|x_i) \\ &= \frac{1 \cdot p_s(b=0)}{\sum_{j=1}^{2} p_s(y=K+1|b=j) p_s(b=j)} p_s(y=K+1|x_i) \\ &= \frac{p_s(b=0)}{p_s(b=0)} \cdot p_s(y=K+1|x_i) = p_s(y=K+1|x_i), \end{aligned} \tag{36}$$

and for $j \in \{1, 2, ..., K\}$ we have:

$$\begin{aligned} p_s(y=j|x_i) &= p_s(y=j|x_i, b=1) p_s(b=1|x_i) + p_s(y=j|x_i, b=0) p_s(b=0|x_i) \\ &= f(x)_j \cdot \rho_s + 0 \cdot (1 - h(x)) = h(x) \cdot f(x_i)_j. \end{aligned} \tag{37}$$

Marginalize Eq. (34) we can also get:

$$p_s(y=j) = \begin{cases} c_j \cdot \rho_s, & j \neq K+1 \\ 1-\rho_s, & j = K+1 \end{cases}, \quad p_t(y=j) = \begin{cases} \pi_j \cdot \rho_t, & j \neq K+1 \\ 1-\rho_t, & j = K+1 \end{cases}, \tag{38}$$

Substituting Eq. (34) and Eq. (38) into the likelihood Eq. (35) we get:

$$\begin{aligned} L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) &= \prod_{i=1}^{N^t} \left(\sum_{j=1}^{K} \frac{h(x) \cdot f(x_i)_j}{\rho_s \cdot c_j} \pi_j \cdot \rho_t + \frac{1-h(x)}{1-\rho_s} \cdot 1 \cdot (1-\rho_t) \right) \cdot Const, \\ &= \prod_{i=1}^{N^t} \left(\frac{\rho_t}{\rho_s} h(x_i) \cdot \sum_{j=1}^{K} \frac{\pi_j}{c_j} f(x_i)_j + \frac{1-\rho_t}{1-\rho_s} \cdot (1-h(x_i)) \right) \cdot Const. \end{aligned} \tag{39}$$

Further substitute Eq. (6) and Eq. (7) into Eq. (39) and then we can get the result.

□

C.4. Proof of Theorem 4.3 (See page 5)

Theorem 4.3. ***(MLE)*** *Under Assumption 3.2, 3.3, the the NLL (5) is convex in $\tilde{\boldsymbol{\pi}}$ (and convex in ρ_t), and the EM algorithm MLE-OLS (Alg. 1) converges to $\boldsymbol{\pi}^{MLE}, \rho_t^{MLE}$ (8).*

Algorithm 5 MLE-OLS

Input: $\mathcal{D}^t_f = \{x^t_i\}_{i=1}^{N^t}, \mathbf{c}, \rho_s, h(x), f(x)$.
Initialize: $\boldsymbol{\pi}^{(0)} \in \Delta^{K-1}_{>0}, \rho_t^{(0)} \in (0,1)$
for $m = 0$ to M **do**
  Construct: $\tilde{\boldsymbol{\pi}}^{(m)}$ based on $\boldsymbol{\pi}^{(m)}, \rho_t^{(m)}$ and Eq. (7).
  E-step: For $j \in (\mathcal{Y} \cup \{K+1\})$, evaluate

$$g_{ij}^{(m)} = \frac{\tilde{\pi}_j^{(m)} / \tilde{c}_j \cdot \tilde{f}(x_i^t)_j}{\sum_{l=1}^{K+1} \tilde{\pi}_l^{(m)} / \tilde{c}_l \cdot \tilde{f}(x_i^t)_l}. \tag{40}$$

  M-step: For $j \in \mathcal{Y}$, evaluate

$$\begin{cases} \pi_j^{(m+1)} = \dfrac{\sum_{i=1}^{N^t} g_{ij}^{(m)}}{N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)}} \\ \rho_t^{(m+1)} = \dfrac{N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)}}{N} \end{cases} \tag{41}$$

end for
Output: $p_t(y = \cdot) = \boldsymbol{\pi}^{(M+1)}, p_t(b=1) = \rho_t^{(M+1)}$.

Proof. **Convexity**: As shown in Lemma 4.2 Eq. (39), the negative log likelihood of $\boldsymbol{\pi}, \rho_t$ given Assumption 3.2, 3.3 and unlabeled target domain dataset $\mathcal{D}^t$ can be written as:

$$-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) = -\sum_{i=1}^{N^t} \log \left(\frac{\rho_t}{\rho_s} h(x_i) \cdot \sum_{j=1}^{K} \frac{\pi_j}{c_j} f(x_i)_j + \frac{1-\rho_t}{1-\rho_s} \cdot (1-h(x_i)) \right) + C. \tag{42}$$

As a function of ρ_t, the NLL can be rewritten as:

$$-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) = -\sum_{i=1}^{N^t} \log(A\rho_t + B) + C, \tag{43}$$

which is a convex function w.r.t. ρ_t.

As a function of $\boldsymbol{\pi}$, the NLL can be rewritten as:

$$-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) = -\sum_{i=1}^{N^t} \log \left(A \sum_{j=1}^{K} \frac{\pi_j}{c_j} f(x_i)_j + B \right) + C, \tag{44}$$

which is a convex function w.r.t $\boldsymbol{\pi}$.

Moreover, same as the close world setting [1], the NLL is convex in the reparameterisation of $\tilde{c}_t$. □

Proof. **EM algorithm**: The NLL objective of MLE defined in Lemma 4.2, Eq. (5) can be rewritten as:

$$-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) = -\sum_{i=1}^{N^t} \log \left(\sum_{j=1}^{K} \frac{\tilde{\pi}_j}{\tilde{c}_j} \tilde{f}(x_i)_j \right) + C, \tag{45}$$

which is reparametrised as the objective of the closed set label shift estimation model MLLS [47] algorithm (Appendix. B, Alg. 3).

As MLE is invariant under reparametrisation [41], and MLLS has been proved to converge to a MLE estimate [1], thus EM algorithm 5 converges to a $\tilde{c}_t^{\text{MLE}}$ and will also converge to a $\boldsymbol{\pi}^{\text{MLE}}, \rho_t^{\text{MLE}}$.

The MLE can be seen as a special case of MAP estimate with prior distribution being 1. In this case, by setting $\boldsymbol{\alpha}^{\textbf{in}} = \mathbf{1}, \alpha_1^{\textbf{out}} = 1, \alpha_2^{\textbf{out}}) = 1$. Proof of EM algorithm for MAP estimate can be found in Proof of Proposition C.2. □

C.5. MAP estimation of target label distribution parameters

MAP estimate: Moreover, if we employ a prior $\boldsymbol{\pi} \sim p(\boldsymbol{\pi}|\boldsymbol{\alpha}^{\textbf{in}})$ over the target label distribution $\boldsymbol{\pi}$, or a prior $\rho_t \sim p(\rho_t|\boldsymbol{\alpha}^{\textbf{out}})$ over the target ID data ratio ρ_t, we can construct the posterior of $\boldsymbol{\pi}$ and ρ_t as:

$$\begin{aligned} -\log p(\boldsymbol{\pi}, \rho_t|\mathcal{D}^t, \boldsymbol{\alpha}) = &-\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) - \log p(\boldsymbol{\pi}|\boldsymbol{\alpha}^{\textbf{in}}) \\ &- \log p(\rho_t|\boldsymbol{\alpha}^{\textbf{out}}) + C. \end{aligned} \tag{46}$$

In this work, inspired by Ye et al. 62, we show that with a Dirichlet prior over $\boldsymbol{\pi} \sim \mathrm{Dir}(K, \boldsymbol{\alpha}^{\textbf{in}})$ or a Beta prior over $\rho_t \sim \mathrm{Beta}(\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}})$, the MAP estimate $\tilde{\boldsymbol{\pi}}^{\text{MAP}}$ can be obtained via another EM algorithm over the objective:

$$\boldsymbol{\pi}^{\text{MAP}}, \rho_t^{\text{MAP}} \in \underset{\tilde{\pi} \in \Delta^K}{\arg\min} - \log p(\boldsymbol{\pi}, \rho_t|\mathcal{D}^t, \boldsymbol{\alpha}), \tag{47}$$

where the details are also provided in Proposition C.2.

Proposition C.2. ***(MAP)*** *Under Assumption 3.2, 3.3, if $\boldsymbol{\pi} \sim Dir(K, \boldsymbol{\alpha}^{in})$ with $\boldsymbol{\alpha}^{in} \in \mathbb{R}^K_{>1}$ and $\rho_t \sim Beta(\alpha_1^{out}, \alpha_2^{out})$ with $\alpha_1^{out}, \alpha_2^{out} \in \mathbb{R}_{>1}$, then*

- *The posterior in Eq. (46) is strictly convex in $\boldsymbol{\pi}$ and strictly convex in ρ_t.*
- *EM algorithm 1 converge to the $\tilde{\boldsymbol{\pi}}^{MAP}$ in Eq. (47).*

Proof. **Convexity**: As shown in the Proof Proposition 4.3, the MLE objective given in Lemma 4.2 is convex on $\boldsymbol{\pi}, \rho_t$.

Since Dirichlet prior $\boldsymbol{\pi} \sim \mathrm{Dir}(K, \boldsymbol{\alpha}^{\textbf{in}})$ with $\boldsymbol{\alpha}^{\textbf{in}} > \mathbf{1}$ is strictly convex on $\boldsymbol{\pi}$. And Beta prior $\rho_t \sim \mathrm{Beta}(\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}})$ with $\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}} > 1$ is strictly convex on ρ_t, the overall posterior:

$$-\log p(\boldsymbol{\pi}, \rho_t|\mathcal{D}^t, \boldsymbol{\alpha}) = -\log L(\boldsymbol{\pi}, \rho_t; \mathcal{D}^t) - \log p(\boldsymbol{\pi}|\boldsymbol{\alpha}^{\textbf{in}}) - \log p(\rho_t|\boldsymbol{\alpha}^{\textbf{out}}) + C \tag{50}$$

is strictly convex on $\boldsymbol{\pi}$ and ρ_t

□

Proof. **EM algorithm:**

Algorithm 6 MAP-OLS

Input: $\mathcal{D}_f^t = \{x_i^t\}_{i=1}^{N^t}$, $\mathbf{c}, \rho_s, h(x), f(x)$, $\boldsymbol{\alpha}^{\textbf{in}}, \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}}$.
Require: $\boldsymbol{\alpha}^{\textbf{in}} \in \mathbb{R}_{>1}^K$, $\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}} \in \mathbb{R}_{>1}$.
Initialize: $\boldsymbol{\pi}^{(0)} \in \Delta_{>0}^{K-1}, \rho_t^{(0)} \in (0,1)$.
Construct: $\tilde{f}$ based on Eq. (6).
for $m = 0$ to M **do**
Construct: $\tilde{\boldsymbol{\pi}}^{(m)}$ based on $\boldsymbol{\pi}^{(m)}, \rho_t^{(m)}$ and Eq. (7).
E-step: For $j \in \mathcal{Y} \cup \{K+1\}$, evaluate

$$g_{ij}^{(m)} = \frac{\tilde{\pi}_j^{(m)}/\tilde{c}_j \cdot \tilde{f}(x_i^t)_j}{\sum_{l=1}^K \tilde{\pi}_l^{(m)}/\tilde{c}_l \cdot \tilde{f}(x_i^t)_l}. \tag{48}$$

M-step: For $j \in \mathcal{Y}$, evaluate

$$\begin{cases} \pi_j^{(m+1)} = \dfrac{\sum_{i=1}^{N^t} g_{ij}^{(m)} + \alpha_j^{\textbf{in}} - 1}{N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)} + \sum_{l=1}^K (\alpha_l^{\textbf{in}} - 1)} \\ \rho_t^{(m+1)} = \dfrac{N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)} + \alpha_1^{\textbf{out}} - 1}{N^t + \alpha_1^{\textbf{out}} + \alpha_2^{\textbf{out}} - 2}. \end{cases} \tag{49}$$

end for
Output: $p_t(y = \cdot) = \boldsymbol{\pi}^{(M+1)}, p_t(b=1) = \rho_t^{(M+1)}$.

To be concise, we will use the notation:

$$\begin{aligned} \tilde{f}(x)_i &= \begin{cases} h(x) \cdot f(x)_i, & i \in \mathcal{Y} \\ 1 - h(x), & i = K+1, \end{cases} \\ \tilde{\boldsymbol{\pi}} &= [\rho_t \cdot \pi_1, ..., \rho_t \cdot \pi_K, 1 - \rho_t]^T \\ \tilde{\mathbf{c}} &= [\rho_s \cdot c_1, ..., \rho_t \cdot c_K, 1 - \rho_s]^T \end{aligned} \tag{51}$$

Remark: We prove the case with the model having both prior $\boldsymbol{\pi} \sim \text{Dir}(K, \boldsymbol{\alpha}^{\textbf{in}})$ and $\rho_t \sim \text{Beta}(\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}})$, where $\boldsymbol{\alpha}^{\textbf{in}} \in \mathbb{R}_{>1}^K$ and $\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}} \in \mathbb{R}_{>1}$. EM algorithms for other cases can be derived similarly by setting $\boldsymbol{\alpha}^{\textbf{in}} = \mathbf{1}$ or $\alpha_1^{\textbf{out}} = 1, \alpha_2^{\textbf{out}} = 1$ or both.

The proof consists of three stages:

1. Identify the latent variable, derive the complete posterior;
2. Construct the $Q(\boldsymbol{\pi}, \rho_t | \boldsymbol{\pi}^{(m)}, \rho_t^{(m)})$ and obtain **E-Step**;
3. Optimize $Q(\boldsymbol{\pi}, \rho_t | \boldsymbol{\pi}^{(m)}, \rho_t^{(m)})$ w.r.t $\boldsymbol{\pi}, \rho_t$ and obtain **M-Step**.

Step 1: As discussed in the main paper (Eq. (5)), we can construct the latent variable $\tilde{Y}_s \sim \text{Cat}(K+1, \tilde{\mathbf{c}})$ and $\tilde{Y}_t \sim \text{Cat}(K+1, \tilde{\boldsymbol{\pi}})$. With $\tilde{Y}_t$ as latent variable, let $\tilde{\mathbb{Y}} = \{\tilde{y}_i^t\}_{i=1}^N$ with $\tilde{y}_i^t \in \mathcal{Y} \cup \{K+1\}$, the complete posterior $p(\boldsymbol{\pi} | \mathcal{D}^t, \tilde{\mathbb{Y}}, \boldsymbol{\alpha}^{\textbf{in}}, \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}})$ can be written as:

$$\begin{aligned} p(\boldsymbol{\pi}, \rho_t | \mathcal{D}_f^t, \tilde{\mathbb{Y}}, \boldsymbol{\alpha}^{\textbf{in}}, \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}}) &= \frac{1}{C} p(\boldsymbol{\pi} | \boldsymbol{\alpha}^{\textbf{in}}) p(\rho_t | \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}}) \prod_{i=1}^N \prod_{j=1}^{K+1} p_t(x_i^t, \tilde{y}_i^t = j; \tilde{\boldsymbol{\pi}}) \\ &= \frac{1}{C} p(\boldsymbol{\pi} | \boldsymbol{\alpha}^{\textbf{in}}) p(\rho_t | \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}}) \prod_{i=1}^N \prod_{j=1}^{K+1} \frac{p_t(\tilde{y}_i^t = j; \tilde{\boldsymbol{\pi}})}{p_s(\tilde{y}_i^t = j)} p_s(\tilde{y}_i^t = j | x_i^t) \\ &= \frac{1}{C} \rho_t^{\alpha_1^{\textbf{out}}} (1 - \rho_t)^{\alpha_2^{\textbf{out}}} \prod_{l=1}^K \pi_l^{\alpha_l^{\textbf{in}} - 1} \prod_{i=1}^N \prod_{j=1}^{K+1} \left(\frac{\tilde{\pi}_j}{\tilde{c}_j} \right)^{\mathbb{I}(\tilde{y}_i^t = j)} \tilde{f}(x_i^t)_j, \end{aligned} \tag{52}$$

where C includes all the terms that are constant w.r.t $\boldsymbol{\pi}, \rho_t$.

Step 2: Given the complete posterior $p(\boldsymbol{\pi}|\mathcal{D}_f^t, \tilde{\mathbb{Y}}, \boldsymbol{\alpha}^{\textbf{in}}, \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}})$, we can construct the $Q(\boldsymbol{\pi}, \rho_t|\boldsymbol{\pi}^{(m)}, \rho_t^{(m)})$ in the **E-Step** as:

$$
\begin{aligned}
Q(\boldsymbol{\pi}, \rho_t|\boldsymbol{\pi}^{(m)}, \rho_t^{(m)}) &= \mathop{\mathbb{E}}_{\tilde{\mathbb{Y}}|\mathcal{D}^t, \boldsymbol{\pi}^{(m)}, \rho_t^{(m)}} \left[\log p(\boldsymbol{\pi}, \rho_t|\mathcal{D}_f^t, \tilde{\mathbb{Y}}, \boldsymbol{\alpha}^{\textbf{in}}, \alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}})\right] \\
&= \mathop{\mathbb{E}}_{\tilde{\mathbb{Y}}|\mathcal{D}^t, \boldsymbol{\pi}^{(m)}, \rho_t^{(m)}} \Bigg[\sum_{i=1}^{N}\sum_{j=1}^{K+1} \mathbb{I}(\tilde{y}_i^t = j)\log\tilde{\pi}_j + \sum_{l=1}^{K}(\alpha_l - 1)\log\pi_l \\
&\qquad + \alpha_1^{\textbf{out}}\cdot\log\rho_t + \alpha_2^{\textbf{out}}\cdot\log(1-\rho_t) + C\Bigg] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{K+1} p_t(\tilde{y}_i^t = j|x_i^t; \tilde{\boldsymbol{\pi}}^{(m)})\log\tilde{\pi}_j + \sum_{l=1}^{K}(\alpha_l - 1)\log\pi_l \\
&\qquad + \alpha_1^{\textbf{out}}\cdot\log\rho_t + \alpha_2^{\textbf{out}}\cdot\log(1-\rho_t) + C \\
&= \sum_{i=1}^{N}\sum_{j=1}^{K+1} g_{ij}^{(m)}\log\tilde{\pi}_j + \sum_{l=1}^{K}(\alpha_l - 1)\log\pi_l \\
&\qquad + \alpha_1^{\textbf{out}}\cdot\log\rho_t + \alpha_2^{\textbf{out}}\cdot\log(1-\rho_t) + C,
\end{aligned}
\tag{53}
$$

where the likelihood $g_{ij}^{(m)} := p_t(\tilde{y}_i^t = j|x_i^t; \boldsymbol{\pi}^{(m)}, \rho^{(m)})$ can be simply obtained via:

$$
g_{ij}^{(m)} = \frac{\frac{\tilde{\pi}_j^{(m)}}{\tilde{c}_j^{(m)}}\tilde{f}(x_i)_j}{\sum_{l=1}^{K+1}\frac{\tilde{\pi}_l^{(m)}}{\tilde{c}_l^{(m)}}\tilde{f}(x_i)_l} \quad \text{for all} \quad j \in \mathcal{Y}\cup\{K+1\}. \tag{54}
$$

Step 3: In the **M-step**, with available $Q(\boldsymbol{\pi}, \rho_t|\boldsymbol{\pi}^{(m)}, \rho_t^{(m)})$, we solve the optimization objective with respect to $\boldsymbol{\pi}$ by fixing ρ_t and vise versa:

$$
\boldsymbol{\pi}^{(m+1)}, \rho_t^{(m+1)} = \mathop{\arg\max}_{\boldsymbol{\pi}\in\Delta^{K-1}, \rho_t\in[0,1]} Q(\boldsymbol{\pi}, \rho_t|\boldsymbol{\pi}^{(m)}, \rho_t^{(m)}) \tag{55}
$$

By substitution, the objective can be rewritten as:

$$
\begin{cases}
\min\limits_{\boldsymbol{\pi}} -\sum\limits_{i=1}^{N^t}\sum\limits_{j=1}^{K+1} g_{ij}^{(m)}\log\tilde{\pi}_j - \sum\limits_{l=1}^{K}(\alpha_l^{\textbf{in}} - 1)\log\pi_l - \alpha_1^{\textbf{out}}\cdot\log\rho_t - \alpha_2^{\textbf{out}}\cdot\log(1-\rho_t) \\
\text{s.t:} \ \sum\limits_{j=1}^{K}\pi_j = 1, \tilde{\boldsymbol{\pi}} = [\rho_t\cdot\pi_1, ..., \rho_t\cdot\pi_K, 1-\rho_t]^T, \\
\quad \pi_i \geq 0 \text{ for } i \in [1, 2, ...K], \rho_t \in [0, 1].
\end{cases}
\tag{56}
$$

Convexity Eq. (56) is just a linear combination of $\log\pi_i$, which is a concave function w.r.t. $\boldsymbol{\pi}$. Knowing that the constraints define a convex set on $\mathbb{R}^K$, therefore Eq. (56) is convex w.r.t $\boldsymbol{\pi}$ and every local minima is a global minima. Similarly, it is also easy to show that Eq. (56) is also convex w.r.t. ρ_t for $\rho_t \in [0, 1]$.

Optimization without inequality constraints With only equality constraints, standard the Lagrangian Multiplier method can be applied. The Lagrangian can be written as:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\pi}, \rho_t, \lambda) &= \sum_{i=1}^{N^t}\sum_{j=1}^{K} g_{ij}^{(m)}\log(\rho_t\cdot\pi_j) + \sum_{i=1}^{N^t} g_{iK+1}^{(m)}\log(1-\rho_t) \\
&\quad + \sum_{j=1}^{K}(\alpha_j^{\textbf{in}} - 1)\log\pi_j + (\alpha_1^{\textbf{out}} - 1)\cdot\log\rho_t + (\alpha_2^{\textbf{out}} - 1)\cdot\log(1-\rho_t) \\
&\quad + \lambda\left(1 - \sum_{j=1}^{K}\pi_j\right).
\end{aligned}
\tag{57}
$$

The optimal $\boldsymbol{\pi}, \rho_t$ can be found by taking all the partial derivative of $\mathcal{L}(\boldsymbol{\pi}, \rho_t, \lambda)$ w.r.t π_j, ρ_t and λ to 0:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \pi_j} = \dfrac{\sum_{i=1}^{N^t} g_{ij}^{(m)}}{\pi_j} + \dfrac{\alpha_j^{\textbf{in}} - 1}{\pi_j} - \lambda = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \rho_t} = \dfrac{\sum_{i=1}^{N^t} \sum_{j=1}^{K} g_{ij}^{(m)} + (\alpha_1^{\textbf{out}} - 1)}{\rho_t} - \dfrac{\sum_{i=1}^{N^t} g_{iK+1}^{(m)} + (\alpha_2^{\textbf{out}} - 1)}{1 - \rho_t} = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{K} \pi_i - 1 = 0. \end{cases} \tag{58}$$

The solution to the above equation set can be written as:

$$\begin{cases} \pi_j = \dfrac{\sum_{i=1}^{N^t} g_{ij}^{(m)} + \alpha_j^{\textbf{in}} - 1}{\lambda} \\ \rho_t = \dfrac{\sum_{i=1}^{N^t} \sum_{j=1}^{K} g_{ij}^{(m)} + \alpha_1^{\textbf{out}} - 1}{N^t + \alpha_1^{\textbf{out}} + \alpha_2^{\textbf{out}} - 2} \\ \lambda = \sum_{i=1}^{N^t} \sum_{j=1}^{K} g_{ij}^{(m)} + \sum_{l=1}^{K} (\alpha_l^{\textbf{in}} - 1). \end{cases} \tag{59}$$

Therefore optimal $\boldsymbol{\pi}, \rho_t$ for $Q(\boldsymbol{\pi}, \rho_t | \boldsymbol{\pi}^{(m)}, \rho_t^{(m)})$ without inequality constraints is given by:

$$\begin{aligned} \pi_j &= \frac{\sum_{i=1}^{N^t} g_{ij}^{(m)} + \alpha_j^{\textbf{in}} - 1}{\sum_{i=1}^{N^t} \sum_{j=1}^{K} g_{ij}^{(m)} + \sum_{l=1}^{K} (\alpha_l^{\textbf{in}} - 1)}, \\ \rho_t &= \frac{\sum_{i=1}^{N^t} \sum_{j=1}^{K} g_{ij}^{(m)} + \alpha_1^{\textbf{out}} - 1}{N^t + \alpha_1^{\textbf{out}} + \alpha_2^{\textbf{out}} - 2} \end{aligned} \tag{60}$$

Proof that the solution satisfies inequality constraints Note that we have:

- $g_{ij}^{(m)}$ in Eq. (54) is non-negative
- $\tilde{c}_i > 0, i = 1, 2...K$ is non-negative
- $\alpha_i^{\textbf{in}} - 1 > 0, i = 1, 2...K$ and $\alpha_1^{\textbf{out}} - 1 > 0, \alpha_2^{\textbf{out}} - 1 > 0$

Therefore we have $\boldsymbol{\pi}^{(t)} > 0 \Rightarrow \boldsymbol{\pi}^{(m+1)} > 0$. Because the optimization problem is convex, when $\pi_j^{(t)} > 0, j = 1, 2, ...K$, Eq. (60) gives the global optimal $\boldsymbol{\pi}^{(m+1)}, \rho_t^{(m+1)}$ for the optimization problem in Eq. (56):

$$\begin{cases} \pi_j^{(m+1)} = \dfrac{\sum_{i=1}^{N^t} g_{ij}^{(m)} + \alpha_j^{\textbf{in}} - 1}{N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)} + \sum_{l=1}^{K} (\alpha_l^{\textbf{in}} - 1)} \quad \text{for all} \quad i \in \mathcal{Y} \\ \rho_t^{(m+1)} = \dfrac{N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)} + \alpha_1^{\textbf{out}} - 1}{N^t + \alpha_1^{\textbf{out}} + \alpha_2^{\textbf{out}} - 2}, \end{cases} \tag{61}$$

given the fact that $\sum_{i=1}^{N^t} \sum_{j=1}^{K} g_{ij}^{(m)} = N^t - \sum_{i=1}^{N^t} g_{iK+1}^{(m)}$.

□

C.6. Proof of Theorem 4.4 (See page 5)

Theorem 4.4. *(**Target ID/OOD ratio correction**) Under Assumption 3.2, 3.3A (without 3.3B), for a classifier $h' : \mathcal{X} \to [0, 1]$ that satisfies (11), given source ID dataset $\mathcal{D}^s$, OOD dataset $\mathcal{D}^o$, target dataset $\mathcal{D}^t$, then for $\delta > 0$, with probability of at least $1 - 2\delta$ we have:*

$$|\rho_t - \hat{\rho}_t^*| \leq \frac{1}{|\mu_1' - \mu_0'|} \sqrt{\frac{2 \log 1/\delta}{\min(|\mathcal{D}^s|, |\mathcal{D}^o|, |\mathcal{D}^t|)}}, \tag{12}$$

where

$$\hat{\rho}_t^* = \frac{\hat{\rho}' - \hat{\mu}_0'}{\hat{\mu}_1' - \hat{\mu}_0'}, \quad \textit{and} \quad \hat{\rho}' := \frac{1}{|\mathcal{D}^t|} \sum_{x_i \in \mathcal{D}^t} h'(x_i), \tag{13}$$

with $\hat{\mu}_1', \hat{\mu}_0'$ and μ_1', μ_0' defined in the same way as (3) *and Theorem. 4.1 but substitute h with h'.*

Proof. For target domain dataset $\mathcal{D}^t$, if we are given ID/OOD label: $\mathcal{D}^t = \mathcal{D}^{\mathbf{ti}} \cup \mathcal{D}^{\mathbf{to}}$, for a practical classifier $h'(x)$ we can write:

$$\begin{aligned} \rho' := \mathbb{E}_{X_t}[h'(x)] &= \mathbb{E}_{X_t|B_t=1}[h'(x)] \cdot p_t(b=1) + \mathbb{E}_{X_t|B_t=0}[h'(x)] \cdot p_t(b=0) \\ &= \rho_t \cdot \mathbb{E}_{X_t|B_t=1}[h'(x)] + (1-\rho_t) \cdot \mathbb{E}_{X_t|B_t=0}[h'(x)] \\ &= \mu_1^t \cdot \rho_t + \mu_0^t \cdot (1-\rho_t), \end{aligned} \tag{62}$$

where:

$$\mu_1^t := \mathbb{E}_{X_t|B_t=1}[h'(x)] \quad \text{and} \quad \mu_0^t := \mathbb{E}_{X_t|B_t=0}[h'(x)]. \tag{63}$$

Rearranging Eq. (62), we have that:

$$\rho_t = \frac{1}{\mu_1^t - \mu_0^t} \rho' - \frac{\mu_0^t}{\mu_1^t - \mu_0^t}, \tag{64}$$

where the equation holds when $\mu_1^t \neq \mu_0^t$:

$$\mu_1^t = \mathbb{E}_{X_t|B_t=1}[h'(x)] \neq \mathbb{E}_{X_t|B_t=0}[h'(x)] = \mu_0^t. \tag{65}$$

Option 1 (11): Under Assumption 3.2, the condition (11) holds implies:

$$\mathbb{E}_{X_s|Y_s=i}[h'(x)] = \mathbb{E}_{X_t|Y_t=j}[h'(x)] \quad \text{for all} \quad i, j \in \mathcal{Y}, \tag{66}$$

then according Eq. (1) we have:

$$\begin{aligned} \mu_1^t = \mathbb{E}_{X_t|B_t=1}[h'(x)] &= \sum_{i=1}^K \mathbb{E}_{X_t|Y_t=i}[p_t(y=i|b=1) \cdot h'(x)] \\ &= \sum_{i=1}^K \mathbb{E}_{X_t|Y_t=i}[\pi_i \cdot h'(x)] = \sum_{i=1}^K \pi_i \cdot \mathbb{E}_{X_t|Y_t=1}[h'(x)] \\ &= \mathbb{E}_{X_t|Y_t=1}[h'(x)] = \mathbb{E}_{X_s|Y_s=1}[h'(x)] \\ &= \sum_{i=1}^K c_i \mathbb{E}_{X_s|Y_s=1}[h'(x)] = \sum_{i=1}^K \mathbb{E}_{X_s|Y_s=i}[c_i \cdot h'(x)] \\ &= \sum_{i=1}^K \mathbb{E}_{X_s|Y_s=i}[p_s(y=i|b=1) \cdot h'(x)] \\ &= \mathbb{E}_{X_t|B_t=1}[h'(x)] = \mu_1'. \end{aligned} \tag{67}$$

Option 2 $\boldsymbol{\pi} = \mathbf{c}$ Condition Eq. (1) can actually be replace with $\boldsymbol{\pi} = \mathbf{c}$ with the results still holds:

$$\begin{aligned} \mu_1^t = \mathbb{E}_{X_t|B_t=1}[h'(x)] &= \sum_{i=1}^K \mathbb{E}_{X_t|Y_t=i}[p_t(y=i|b=1) \cdot h'(x)] \\ &= \sum_{i=1}^K \mathbb{E}_{X_t|Y_t=i}[\pi_i \cdot h'(x)] = \sum_{i=1}^K \mathbb{E}_{X_s|Y_s=i}[c_i \cdot h'(x)] \\ &= \sum_{i=1}^K \mathbb{E}_{X_s|Y_s=i}[p_s(y=i|b=1) \cdot h'(x)] = \mathbb{E}_{X_t|B_t=1}[h'(x)] = \mu_1', \end{aligned} \tag{68}$$

where $\mu_1' := \mathbb{E}_{X_s|B_s=1}[h'(x)]$.

For both options we have:

$$\begin{aligned}\mu_0^t = \mathbb{E}_{X_t|B_t=0}[h'(x)] &= \mathbb{E}_{X_t|Y_t=K+1}[p_t(y=K+1|b=0)\cdot h'(x)] \\ &= \mathbb{E}_{X_s|Y_s=K+1}[p_s(y=K+1|b=0)\cdot h'(x)] = \mathbb{E}_{X_s|B_s=0}[h'(x)] = \mu_0'.\end{aligned} \tag{69}$$

where $\mu_0' := \mathbb{E}_{X_s|B_s=0}[h'(x)]$ are defined in the same way as μ_1, μ_0 defined in Theorem 4.1 but substitute h as h'.

The expectations can be approximated by $\hat{\mu}_1', \hat{\mu}_1'$ with source domain ID dataset $\mathcal{D}^s$ and OOD dataset $\mathcal{D}^{\mathbf{o}}$ (Eq. (3)). Moreover, $\mathbb{E}_{X_t}[h(x)] = \rho$ can be estimated with $\hat{\rho}$ given target dataset $\mathcal{D}^t$:

$$\hat{\mu}_1' := \frac{1}{|\mathcal{D}^s|}\sum_{x_i\in\mathcal{D}^s} h'(x_i), \quad \hat{\mu}_0' := \frac{1}{|\mathcal{D}^{\mathbf{o}}|}\sum_{x_i\in\mathcal{D}^{\mathbf{o}}} h'(x_i) \quad \text{and} \quad \hat{\rho}' := \frac{1}{|\mathcal{D}^t|}\sum_{x_i\in\mathcal{D}^t} h'(x_i). \tag{70}$$

Therefore as long as (11) holds, we can use $\mathcal{D}^s$ and $\mathcal{D}^{\mathbf{o}}$ to estimate ρ_t with Eq. (64):

$$\rho_t \approx \hat{\rho}_t^* := \frac{1}{\hat{\mu}_1' - \hat{\mu}_0'}\hat{\rho}' - \frac{\hat{\mu}_0'}{\hat{\mu}_1' - \hat{\mu}_0'}, \tag{71}$$

Note that since $h'(x) \in [0,1]$, Hoeffding's inequality guarantees for some small $\epsilon > 0$:

$$\begin{aligned} p\left(|\mu_0' - \hat{\mu}_0'| \geq \epsilon\right) &\leq 2e^{-2|\mathcal{D}^{\mathbf{so}}|\epsilon^2} \\ p\left(|\mu_1' - \hat{\mu}_1'| \geq \epsilon\right) &\leq 2e^{-2|\mathcal{D}^{\mathbf{si}}|\epsilon^2} \\ p\left(|\rho' - \hat{\rho}'| \geq \epsilon\right) &\leq 2e^{-2|\mathcal{D}^t|\epsilon^2}, \end{aligned} \tag{72}$$

Therefore with high probability of at least $1 - 2e^{-2\min(|\mathcal{D}^{\mathbf{si}}|, |\mathcal{D}^{\mathbf{so}}|, |\mathcal{D}^t|)\epsilon^2}$ we have:

$$\begin{cases} \rho_t - \hat{\rho}_t^* \leq \dfrac{|\rho' - \mu_0'|}{|\mu_1' - \mu_0'|} - \dfrac{|\rho' - \epsilon - \mu_0' - \epsilon|}{|\mu_1' + \epsilon - \mu_0' - \epsilon|} = \dfrac{2\epsilon}{|\mu_1' - \mu_0'|} \\ \rho_t - \hat{\rho}_t^* \geq \dfrac{-|\rho' - \mu_0'|}{|\mu_1' - \mu_0'|} - \dfrac{-|\rho' + \epsilon - \mu_0' + \epsilon|}{|\mu_1' - \epsilon - \mu_0' + \epsilon|} = \dfrac{-2\epsilon}{|\mu_1' - \mu_0'|}, \end{cases} \tag{73}$$

which is equivalent to:

$$|\rho_t - \hat{\rho}_t^*| \leq \frac{2\epsilon}{|\mu_1' - \mu_0'|}. \tag{74}$$

Letting $\delta := e^{-2\min(|\mathcal{D}^{\mathbf{si}}|, |\mathcal{D}^{\mathbf{so}}|, |\mathcal{D}^t|)\epsilon^2}$, rearrange the equations and we get the result.

□

C.7. Further Discussion on ρ_t correction model

This section further discusses the ρ_t correction model Eq. (14) proposed in §4.4 in our main paper. The model adjusts ρ_t^{MLE} and ρ_t^{MAP} obtained in Alg. 1 with Eq. (14), which is based on Theorem 4.4.

We will show that for a special case, the MLE ρ_t^{MLE} defined in MLE objective Eq. (8) will have a closed-form solution, which is simply averaging the response of $h(x)$ on target dataset $\mathcal{D}^t$:

Lemma C.3. *Under Assumption 3.2,3.3, if $\boldsymbol{\pi} = \mathbf{c}$ and $h : \mathcal{X} \to \{0,1\}$, then the ρ_t^{MLE} defined in Eq. (8) can be obtained given target dataset $\mathcal{D}^t$ via:*

$$\rho_t^{MLE} = \frac{1}{N^t}\sum_{i=1}^{N^t} h(x_i). \tag{75}$$

Proof. When Assumption 3.3B is satisfied, given the information available, substituting:

$$p_s(b=1|x) = h(x) \in \{0,1\} \quad \text{and} \quad \boldsymbol{\pi} = \mathbf{c} \tag{76}$$

into the NLL in Eq. (5) and we have:

$$
\begin{aligned}
-\log L(\rho_t;\mathcal{D}^t) &= -\sum_{i=1}^{N^t}\log\left(\frac{\rho_t}{\rho_s}h(x_i)\cdot\sum_{j=1}^{K}\frac{\pi_j}{c_j}f(x_i)_j+\frac{1-\rho_t}{1-\rho_s}\cdot(1-h(x_i))\right)+C \\
&= -\sum_{i=1}^{N^t}\log\left(\frac{\rho_t}{\rho_s}h(x_i)\cdot 1+\frac{1-\rho_t}{1-\rho_s}\cdot(1-h(x_i))\right)+C \\
&= -\sum_{i=1}^{N^t}\mathbb{I}_1(h(x_i))\cdot\log\left(\frac{\rho_t}{\rho_s}\right)-\sum_{i=1}^{N^t}(1-\mathbb{I}_1(h(x_i)))\cdot\log\left(\frac{1-\rho_t}{1-\rho_s}\right)+C
\end{aligned}
\tag{77}
$$

Let the derivative w.r.t. ρ_t equals 0 and we have:

$$
\begin{aligned}
\frac{d\left(-\log L(\rho_t;\mathcal{D}^t)\right)}{d\rho_t} &= -\sum_{i=1}^{N^t}\mathbb{I}_1(h(x_i))\cdot\frac{\rho_s}{\rho_t}\cdot\frac{1}{\rho_s}-\sum_{i=1}^{N^t}(1-\mathbb{I}_1(h(x_i)))\cdot\frac{1-\rho_s}{1-\rho_t}\cdot\frac{-1}{1-\rho_s} \\
&= -\sum_{i=1}^{N^t}\mathbb{I}_1(h(x_i))\cdot\frac{1}{\rho_t}+\sum_{i=1}^{N^t}(1-\mathbb{I}_1(h(x_i)))\cdot\frac{1}{1-\rho_t}=0
\end{aligned}
\tag{78}
$$

Solve the above equation for ρ_t and we get:

$$
\rho_t=\frac{1}{N^t}\sum_{i=1}^{N^t}h(x_i),
\tag{79}
$$

which is the closed-form solution to the MLE objective Eq. (5) under the special setting of no ID label shift ($\boldsymbol{\pi}=\mathbf{c}$) and a discrete ID/OOD classifier ($h:\mathcal{X}\to\{0,1\}$).

□

As shown in Lemma C.3, when $\boldsymbol{\pi}=\mathbf{c}$ and $h:\mathcal{X}\to\{0,1\}$, ρ_t^{MLE} can be obtained by averaging $h(x)=p_s(b=1|x)$ over the target dataset $\mathcal{D}^t$ based on Assumption 3.3B, i.e. $h(x)\neq p_s(b=1|x)$ when the assumption is not satisfied, according the proof of Theorem 4.4, the condition $\boldsymbol{\pi}=\mathbf{c}$ enable us to use:

$$
\hat{\rho}_t^*=\frac{\hat{\rho}-\hat{\mu}_0'}{\hat{\mu}_1'-\hat{\mu}_0'},\quad\text{and}\quad\hat{\rho}':=\frac{1}{|\mathcal{D}^t|}\sum_{x_i\in\mathcal{D}^t}h(x_i),
\tag{80}
$$

to obtain the estimate of ground truth ρ_t.

D. Experimental Setup

D.1. ID Classifier Details

Our code for training the ID classifier and constructing the OOD classifier is mainly based on the open source project OpenOOD [61, 64] on OOD detection. The project is publicly available at `https://github.com/Jingkang50/OpenOOD`.

We follow the basic setup in OpenOOD to train the ID classifier f, where we train a ResNet18 model for the CIFAR10/100 and ImageNet-200 datasets. Each model is trained 3 times with different random seeds.

Dataset	Model	Setup	optimizer	lr	weight decay	epoch
CIFAR10	ResNet18	Train from Scratch	SGD	0.1	$5e^{-4}$	100
CIFAR100	ResNet18	Train from Scratch	SGD	0.1	$5e^{-4}$	100
ImageNet-200	ResNet18	Train from Scratch	SGD	0.1	$5e^{-4}$	90

Table 6. Source domain ID classifier f setup used in our model. The setup follows exactly the OpenOOD project implementation (retrieved on May 2024).

D.2. OOD Classifier Details

We use the implementation provided in OpenOOD project to construct the OOD detection binary classifiers h proposed by OpenMax [4], Ash [8], MLS [22], ReAct [50] and KNN [51]. All the OOD detection models are post-hoc inference models based on the ID classifier f. The detailed hyper-parameter setups of each OOD detector are listed in Tab. 8, where:

- OpenMax has no official implementations, we follow the OpenOOD implementation with the same hyperparameter provided by the code.
- KNN follows the exact hyperparameter setup of the original paper K=50, which is also adopted in the OpenOOD.
- MLS does not require any hyperparameter.
- ASH has one hyperparameter "percentile", which is obtained with a parameter search among [65, 70, 75, 80, 85, 90, 95] over a validation set (subset of the source domain dataset $\mathcal{D}^s$) provided by OpenOOD, the original paper simply choose 65.
- ReAct has one hyperparameter "percentile", which is also obtained with a parameter search among [85, 90, 95, 99] over a subset of $\mathcal{D}^s$ provided by OpenOOD, the original simply chose 90.

Model Name	Source Code	Date of Retrieval
OpenOOD [61]	`https://github.com/Jingkang50/OpenOOD`	May 2024
OpenMax [4]	`https://github.com/Jingkang50/OpenOOD`	May 2024
KNN [51]	`https://github.com/deeplearning-wisc/knn-ood`	May 2024
MLS [22]	`https://github.com/Jingkang50/OpenOOD`	May 2004
Ash [8]	`https://github.com/andrijazz/ash`	May 2024
ReAct [50]	`https://github.com/deeplearning-wisc/react`	May 2024

Table 7. Source code details of reproduced OOD detection models. The code for OpenMax, KNN, MLS, Ash and ReAct have been collected in the OpenOOD project and can be directly tested within the project.

OOD classifier	hyper-parameters
OpenMax	Weibull fitting: alpha=3, threshold=0.9, tail=20; coreset_sampling_ratio=0.01;
KNN	# of nearest neighbor $K = 50$
MLS	-
Ash	parameter search on percentile=[65, 70, 75, 80, 85, 90, 95]
ReAct	parameter search on percentile=[85, 90, 95, 99]

Table 8. Detailed hyper-parameter setups of the OOD detectors used in our work. All the hyper-parameter setup are following the default setups provided by the OpenOOD project [61] (retrieved on May 2024).

Output Re-scaling: Existing OOD classifiers focus more on ID/OOD separation and hence usually output a real valued scalar instead of $[0, 1]$ confidence. For example, the MLS [22] model actually outputs the max logit of the ID classifier's prediction. To use these OOD classifiers in our OSLS estimation model, we need to re-scale the output of these OOD classifier in the binary range $[0, 1]$.

In this work, we re-scale an OOD classifier $h' : \mathcal{X} \to \mathbb{R}$ to a binary classifier $h : \mathcal{X} \to [0, 1]$ with two approaches: **logistic regression** and **thresholding**. The logistic regression model $h_0 : \mathbb{R}_+ \to [0, 1]$ is trained based on the source domain ID dataset $\mathcal{D}^s$ and reference OOD dataset $\mathcal{D}^{\mathbf{o}}$ (see Fig. 2). On the other hand, the thresholding approach obtains the threshold by computing the median values of the output of OOD classifier h' given ID dataset $\mathcal{D}^s$ and OOD dataset $\mathcal{D}^{\mathbf{o}}$. The threshold is picked as the average of the two median values. Details or the two re-scaling models are described in Tab. 9.

Dataset	Re-scaling Model	Model Setup
CIFAR10/100	Logistic Regression	epoch 100; optimizer: SGD; batch_size: 512; lr 0.05; lr_scheduler: Cosine; loss: BCE; $h(x) = 1/(1 + e^{-w \cdot h'(x)+b})$
ImageNet-200	Thresholding	$h(x) = \begin{cases} 1, & h'(x) > (\text{median}(h'(\mathcal{D}^s)) + \text{median}(h'(\mathcal{D}^{\mathbf{o}})))/2 \\ 0, & \text{Otherwise} \end{cases}.$

Table 9. Re-scaling model setup that normalize the output of a OOD classifier into $[0, 1]$.

Although thresholding approach only outputs $\{0, 1\}$ instead of a continuous confidence, it is suitable for our ρ_t correction model (Sec. 4.4) because the linear correction approach has theoretical guarantees when the ID/OOD classifier $h(x)$ outputs binary values (Theorem 4.4). Moreover, OOD detectors on large-scale datasets are more likely to violate Assumption 3.3, thus the ρ_t correction model might become more necessary.

D.3. OOD reference Dataset details

As discussed in the main paper, the reference OOD dataset $\mathcal{D}^{\mathbf{o}}$ is generated by linear combination of Gaussian noise and ground truth samples in source domain ID dataset $\mathcal{D}^s$. The hyper-parameters used γ, T used in the OOD dataset generation process and $\hat{\mu}_0$ rescaling are: CIFAR10: $\gamma = 0.2, T = 2$, CIFAR100: $\gamma = 0.1, T = 2$, ImageNet-200: $\gamma = 0.2, T = 2$.

D.4. Datasets Details

Train Datasets: We use the standard CIFAR10/100 and ImageNet-200 datasets as ID datasets, with the detailed information provided in Tab. 10.

Dataset	Train # samples	Val # samples	Test # samples	# of Classes
CIFAR10	50k	9000	1000	10
CIFAR100	50k	9000	1000	100
ImageNet-200	260k	1000	9000	200

Table 10. Detailed information of ID datasets.

Test Datasets: We use the OOD datasets setup provided in the OpenOOD project, where validation sets of CIFAR10 and CIFAR100 are used as OOD datasets, with each contains 9000 samples. For the other OOD datatsets, TinyImageNet has 7793 samples, MNIST has 70000 samples, SVHN has 26032 samples, Texture has 5640 samples, Places has 35195 samples, SSB has 49000 samples, NINCO has 5879 sampels, iNaturalist has 10000 samples, OpenImage-O has 15869 samples. Many of these OOD datasets are actually subsampled from the original datasets to avoid overlapping in classes.

Test label shifts: We follow the MAPLS [62] official code to adjust the test datasets for ordered Long-Tail and Dirichlet label shift (retrieved in May 2024). More detailed information for different test set shifts can be found in the following Tab. 11.

Label Shift	Shift Parameters	OOD/ID data ratio r
Dirichlet	$\alpha = 1.0, 10.0$ (2500 samples)	$r = 1, 0.1, 0.01$
Ordered LT	$100, 50, 10$ "Forward/Backward"	$r = 1, 0.1, 0.01$

Table 11. **Types of label shift in our experiment,** including Dirichlet shift with different shift parameter α and Ordered Long-Tailed (LT) shift with different imbalance factors under forward and backward order.

D.5. Closed Set Label Shift Estimation Model details

We test closed set label shift estimation model BBSE [32], MLLS [47], RLLS [2] and MAPLS [62] based on the official implementation of MAPLS provided in `https://github.com/ChangkunYe/MAPLS` retrieved on June 2024. These models are used to test on the open set label shift dataset without any adjustment of hyper-parameters or other setups.

According to MAPLS, MLLS code is provided by Alexandari *et al*. [1] which has included the source code of RLLS [2] and BBSE [32] with their original github page provided in the Tab. 12. Only RLLS has a hyperparameter in their model. We follow Alexandari *et al*. [1] and the RLLS original implementation to set the hyperparameter to be $\alpha = 0.01$.

Model Name	Source Code	Date of Retrieval
MLLS [1, 47]	https://github.com/kundajelab/labelshiftexperiments https://github.com/kundajelab/abstention	Aug 2022 Aug 2022
BBSE [32]	https://github.com/flaviovdf/label-shift	Aug 2022
RLLS [2]	https://github.com/Angie-Liu/labelshift	Aug 2022

Table 12. Source Code details of reproduced existing label shift estimation models.

D.6. EM algorithm

We use the same EM algorithm running procedure as MAPLS [62] proposed in the closed set label shift problem. Specifically, the procedure is as follows: 1) Initialize the target label distribution to be the same as source label distribution, i.e. $\boldsymbol{\pi}^{(0)} = \hat{\mathbf{c}}$ and $\rho_t^{(0)} = \hat{\rho}_s$, 2) Run EM algorithm 1 for 100 epoch to ensure convergence and 3) Output $\boldsymbol{\pi}^{(101)}$ and $\boldsymbol{\pi}^{(101)}$.

For the MAP estimate, we use the Adaptive Prior Learning (APL) model proposed by MAPLS [62] to determine parameter $\boldsymbol{\alpha}^{\textbf{in}} \in \mathbb{R}^K_{>1}$ in the Dirichlet prior for ID classes and use no Bernoulli prior ($\alpha_1^{\textbf{out}}, \alpha_2^{\textbf{out}} = \mathbf{1}$).

E. More Visualizations and Ablation studies

E.1. Target ρ_t Estimation

The full experiment visualizations of MAP estimate $\hat{\rho}_t^*$ with our model, $\hat{\rho}_t$ with our model but no ρ_t correction in Sec. 4.4 and ground truth ρ_t on different ID/OOD datasets with different label shift are listed in the following figures.

- Fig. 5 includes results on Dirichlet ID label shift on CIFAR10/100 datasets.
- Fig. 6 includes results on LT10 ID label shift on CIFAR10/100 datasets.
- Fig. 7 includes results on LT50 ID label shift on CIFAR10/1000 datasets.
- Fig. 8 includes results on LT100 ID label shift on CIFAR10/1000 datasets.
- Fig. 9 includes results on LT10/LT100 ID label shift on ImageNet-200 dataset.

As seen in the figures, the estimation result of our model exhibits a linear correlation with the ground truth, which is explained by our analysis in Theorem 4.4. Moreover, our ρ_t correction model is able to adjust the predicted $\hat{\rho}_t$ to $\hat{\rho}_t^*$ that is closer to the ground truth.

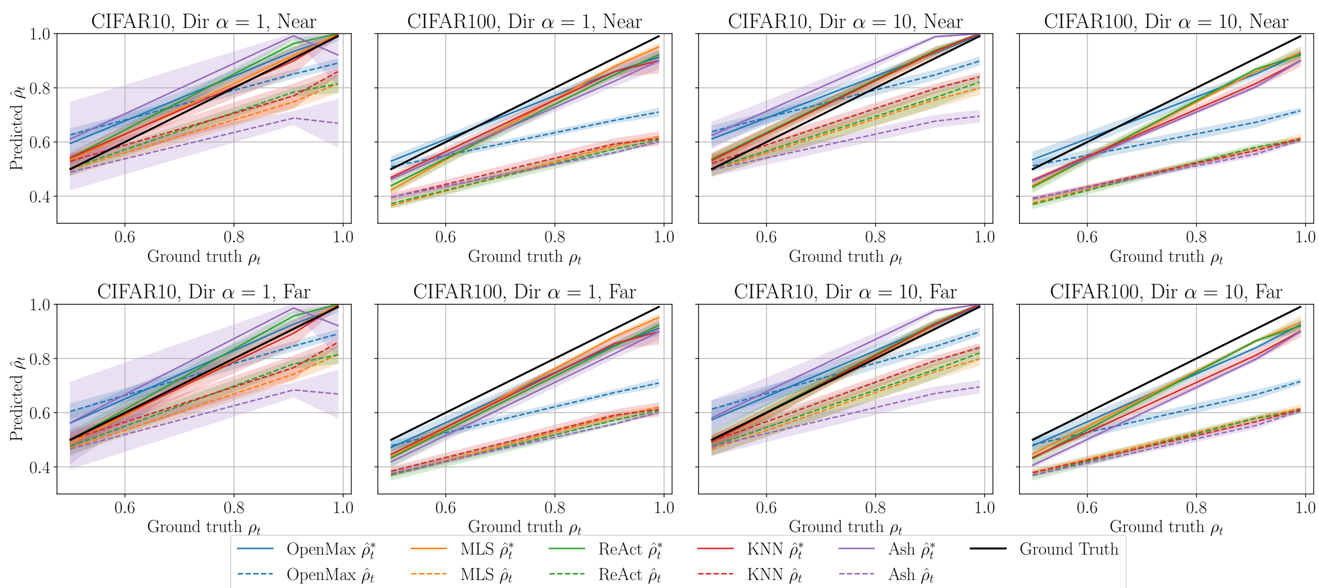

Figure 5. **Estimation result comparison of $\hat{\rho}_t^*$ by our model (Solid lines), $\hat{\rho}_t$ by our model but without ρ_t correction (Sec. 4.4) (Dashed lines) based on different OOD classifiers and the Ground truth ρ_t (Black, Solid line), on CIFAR10/100 dataset with Dirichlet shift and Near + Far OOD dataset (Tab. 1).** Shaded area are $\pm$ one standard deviation over corresponding OOD datasets and three independent ID classifiers.

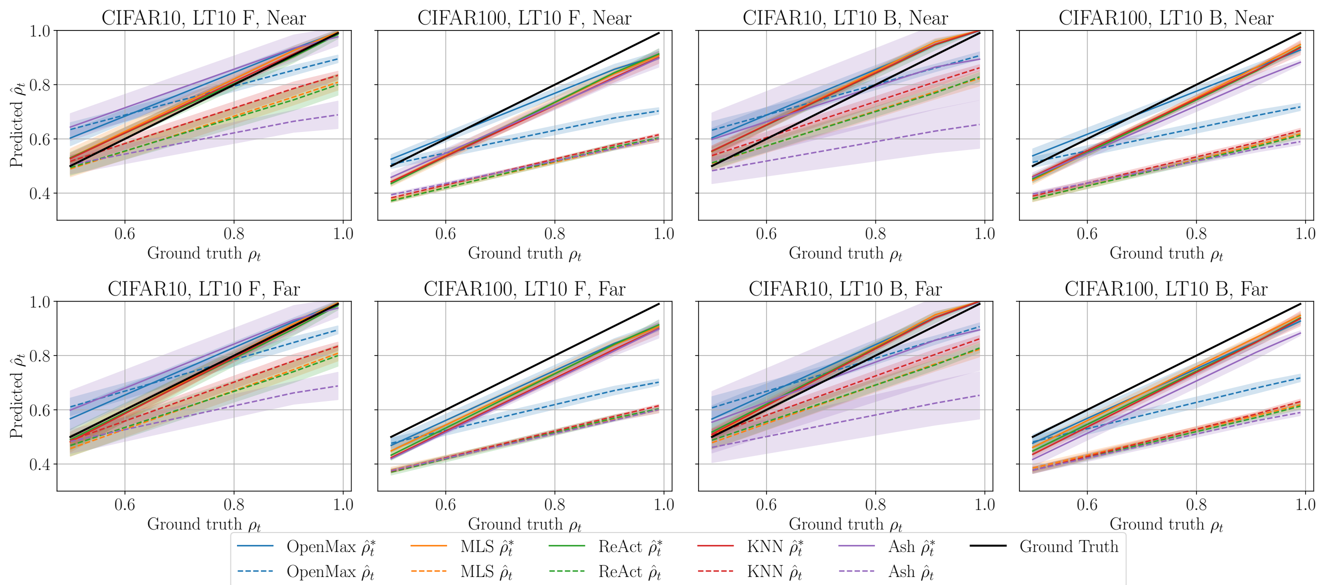

Figure 6. **Estimation result comparison of $\hat{\rho}_t^*$ by our model (Solid lines), $\hat{\rho}_t$ by our model but without ρ_t correction (Sec. 4.4) (Dashed lines) based on different OOD classifiers and the Ground truth ρ_t (Black, Solid line), on CIFAR10/100 dataset with LT10 shift ("F" for Forward and "B" for Backward) and Near + Far OOD dataset (Tab. 1).** Shaded area are $\pm$ one standard deviation over corresponding OOD datasets and three independent ID classifiers.

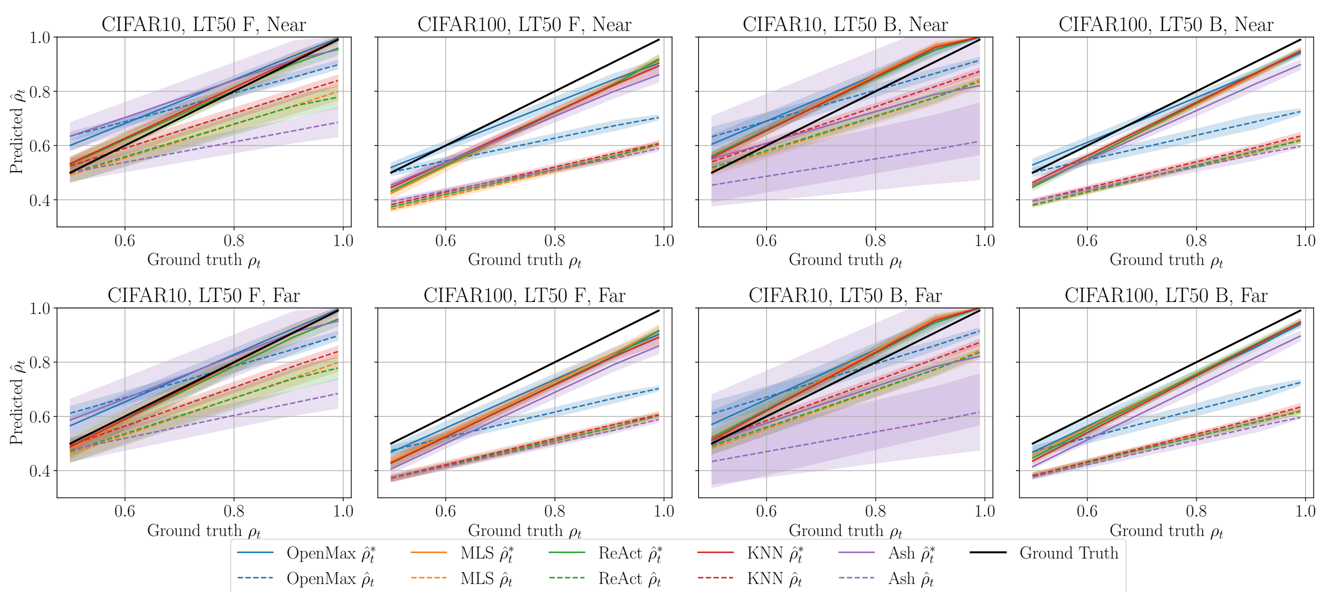

Figure 7. **Estimation result comparison of $\hat{\rho}_t^*$ by our model (Solid lines), $\hat{\rho}_t$ by our model but without ρ_t correction (Sec. 4.4) (Dashed lines) based on different OOD classifiers and the Ground truth ρ_t (Black, Solid line), on the CIFAR10/100 dataset with LT50 shift ("F" for Forward and "B" for Backward) and Near + Far OOD dataset (Tab. 1).** Shaded areas are $\pm$ one standard deviation over corresponding OOD datasets and three independent ID classifiers.

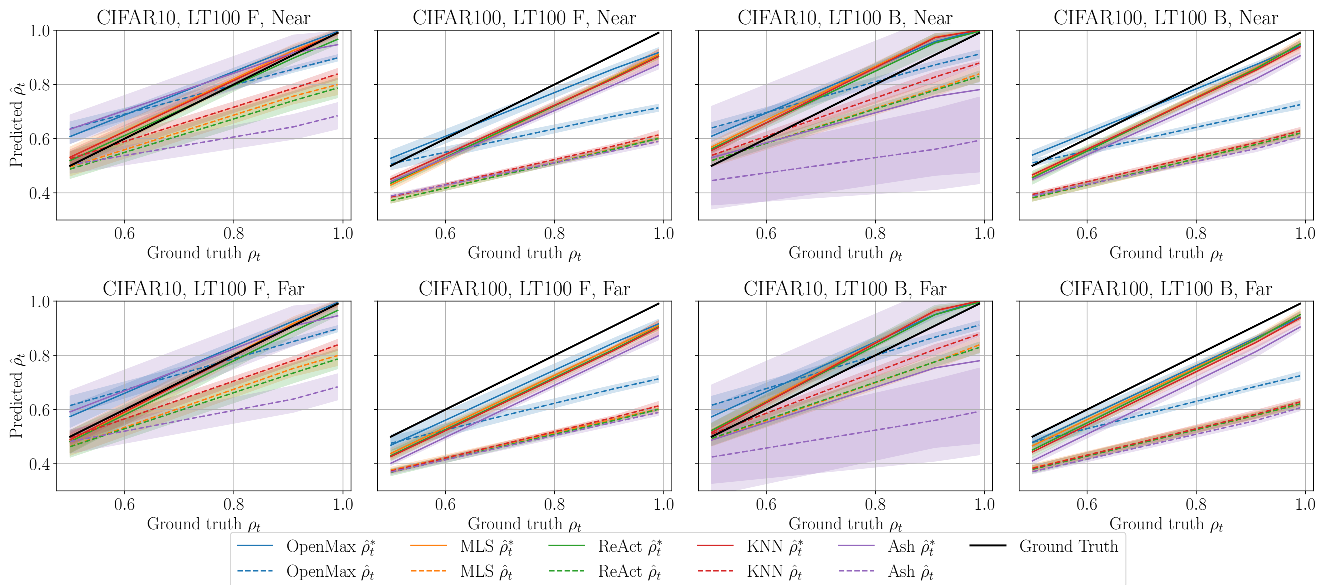

Figure 8. **Estimation result comparison of $\hat{\rho}_t^*$ by our model (Solid lines), $\hat{\rho}_t$ by our model but without ρ_t correction (Sec. 4.4) (Dashed lines) based on different OOD classifiers and the Ground truth ρ_t (Black, Solid line), on the CIFAR10/100 dataset with LT100 shift ("F" for Forward and "B" for Backward) and Near + Far OOD dataset (Tab. 1).** Shaded area are $\pm$ one standard deviation over corresponding OOD datasets and three independent ID classifiers.

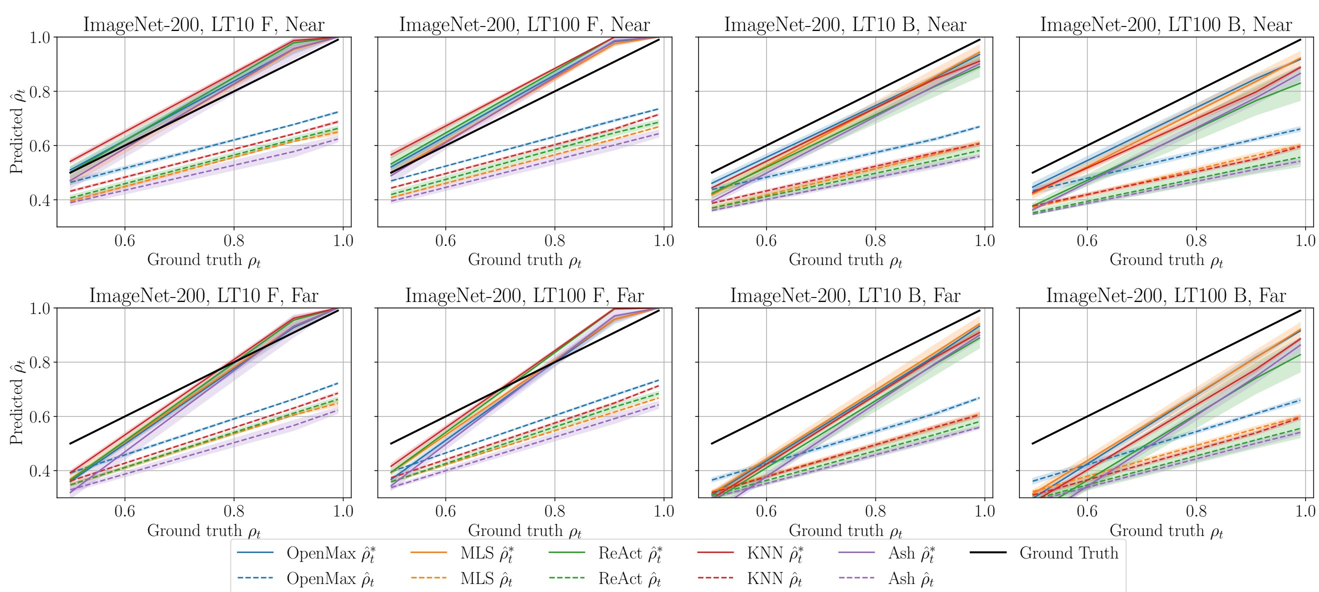

Figure 9. **Estimation result comparison of $\hat{\rho}_t^*$ by our model (Solid lines), $\hat{\rho}_t$ by our model but without ρ_t correction (Sec. 4.4) (Dashed lines) based on different OOD classifiers and the Ground truth ρ_t (Black, Solid line), on the ImageNet-200 dataset with LT10/LT100 shift ("F" for Forward and "B" for Backward) and Near + Far OOD dataset (Tab. 1).** Shaded area are $\pm$ one standard deviation over corresponding OOD datasets and three independent ID classifiers.

E.2. Hyperparameter sensitivity ablation

This section provides the ablation study of the sensitivity of the hyperparameter γ in Eq. (15) when generating pseudo OOD samples with Gaussian noise:

$$\mathcal{D}^{\mathbf{o}}_{\gamma} = \{(1-\gamma)\cdot x_i + \gamma \cdot \epsilon | x_i \in \mathcal{D}^s, \epsilon \sim \mathcal{N}(0,1)\}. \tag{81}$$

As shown in Tab. 13, our model exhibits stable performance when γ varies.

Dataset			CIFAR100								
ID label Shift param			LT10 Forward			LT50 Forward			LT100 Forward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.560_{\pm0.038}$	$0.121_{\pm0.027}$	$0.107_{\pm0.026}$	$0.758_{\pm0.057}$	$0.171_{\pm0.044}$	$0.136_{\pm0.031}$	$0.841_{\pm0.038}$	$0.188_{\pm0.048}$	$0.151_{\pm0.030}$
		Far	$4.128_{\pm0.245}$	$0.253_{\pm0.028}$	$0.109_{\pm0.027}$	$4.370_{\pm0.301}$	$0.291_{\pm0.043}$	$0.139_{\pm0.034}$	$4.431_{\pm0.228}$	$0.306_{\pm0.054}$	$0.153_{\pm0.031}$
MLLS		Near	$0.906_{\pm0.061}$	$0.114_{\pm0.028}$	$0.088_{\pm0.028}$	$1.029_{\pm0.066}$	$0.155_{\pm0.044}$	$0.105_{\pm0.028}$	$1.072_{\pm0.067}$	$0.150_{\pm0.042}$	$0.113_{\pm0.034}$
		Far	$9.633_{\pm1.442}$	$0.348_{\pm0.057}$	$0.092_{\pm0.028}$	$9.910_{\pm1.551}$	$0.380_{\pm0.044}$	$0.108_{\pm0.030}$	$9.896_{\pm1.523}$	$0.373_{\pm0.065}$	$0.115_{\pm0.035}$
RLLS		Near	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.404_{\pm0.000}$	$1.402_{\pm0.000}$	$1.402_{\pm0.000}$
		Far	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.404_{\pm0.000}$	$1.403_{\pm0.000}$	$1.402_{\pm0.000}$
MAPLS		Near	$0.700_{\pm0.034}$	$0.114_{\pm0.018}$	$0.091_{\pm0.019}$	$0.878_{\pm0.037}$	$0.164_{\pm0.033}$	$0.120_{\pm0.019}$	$0.946_{\pm0.041}$	$0.175_{\pm0.031}$	$0.135_{\pm0.026}$
		Far	$7.469_{\pm1.122}$	$0.290_{\pm0.040}$	$0.094_{\pm0.018}$	$7.758_{\pm1.196}$	$0.340_{\pm0.029}$	$0.123_{\pm0.020}$	$7.779_{\pm1.171}$	$0.350_{\pm0.046}$	$0.138_{\pm0.026}$
Open Set Label Shift estimation models											
Baseline			$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$
ours	$\gamma=0.1$	Near	$\mathbf{0.396}_{\pm0.018}$	$\mathbf{0.043}_{\pm0.004}$	$\mathbf{0.046}_{\pm0.009}$	$\mathbf{0.470}_{\pm0.027}$	$\mathbf{0.068}_{\pm0.013}$	$\mathbf{0.073}_{\pm0.013}$	$\mathbf{0.508}_{\pm0.005}$	$\mathbf{0.085}_{\pm0.017}$	$\mathbf{0.078}_{\pm0.015}$
		Far	$2.152_{\pm0.396}$	$\mathbf{0.082}_{\pm0.008}$	$\mathbf{0.047}_{\pm0.009}$	$2.224_{\pm0.323}$	$\mathbf{0.104}_{\pm0.010}$	$\mathbf{0.074}_{\pm0.013}$	$2.426_{\pm0.329}$	$\mathbf{0.118}_{\pm0.012}$	$\mathbf{0.079}_{\pm0.015}$
	$\gamma=0.2$	Near	$0.473_{\pm0.008}$	$\mathbf{0.039}_{\pm0.005}$	$\mathbf{0.032}_{\pm0.001}$	$\mathbf{0.573}_{\pm0.023}$	$\mathbf{0.062}_{\pm0.002}$	$\mathbf{0.051}_{\pm0.004}$	$\mathbf{0.589}_{\pm0.013}$	$\mathbf{0.063}_{\pm0.008}$	$\mathbf{0.057}_{\pm0.005}$
		Far	$3.100_{\pm0.115}$	$\mathbf{0.094}_{\pm0.007}$	$\mathbf{0.034}_{\pm0.001}$	$3.077_{\pm0.221}$	$\mathbf{0.111}_{\pm0.011}$	$\mathbf{0.053}_{\pm0.004}$	$3.144_{\pm0.221}$	$\mathbf{0.118}_{\pm0.009}$	$\mathbf{0.058}_{\pm0.005}$
	$\gamma=0.3$	Near	$0.480_{\pm0.031}$	$\mathbf{0.034}_{\pm0.002}$	$\mathbf{0.034}_{\pm0.002}$	$\mathbf{0.543}_{\pm0.022}$	$\mathbf{0.062}_{\pm0.008}$	$\mathbf{0.055}_{\pm0.008}$	$\mathbf{0.601}_{\pm0.016}$	$\mathbf{0.057}_{\pm0.008}$	$\mathbf{0.056}_{\pm0.011}$
		Far	$3.069_{\pm0.082}$	$\mathbf{0.089}_{\pm0.004}$	$\mathbf{0.036}_{\pm0.002}$	$3.268_{\pm0.238}$	$\mathbf{0.116}_{\pm0.010}$	$\mathbf{0.057}_{\pm0.008}$	$3.236_{\pm0.104}$	$\mathbf{0.114}_{\pm0.008}$	$\mathbf{0.057}_{\pm0.012}$
	$\gamma=0.4$	Near	$0.482_{\pm0.039}$	$\mathbf{0.035}_{\pm0.002}$	$\mathbf{0.032}_{\pm0.004}$	$\mathbf{0.571}_{\pm0.027}$	$\mathbf{0.067}_{\pm0.012}$	$\mathbf{0.056}_{\pm0.010}$	$\mathbf{0.609}_{\pm0.027}$	$\mathbf{0.062}_{\pm0.009}$	$\mathbf{0.053}_{\pm0.003}$
		Far	$3.093_{\pm0.068}$	$\mathbf{0.100}_{\pm0.007}$	$\mathbf{0.034}_{\pm0.004}$	$3.272_{\pm0.117}$	$\mathbf{0.123}_{\pm0.010}$	$\mathbf{0.057}_{\pm0.010}$	$3.314_{\pm0.076}$	$\mathbf{0.123}_{\pm0.002}$	$\mathbf{0.054}_{\pm0.002}$
	$\gamma=0.5$	Near	$0.486_{\pm0.025}$	$\mathbf{0.041}_{\pm0.003}$	$\mathbf{0.032}_{\pm0.004}$	$\mathbf{0.578}_{\pm0.017}$	$\mathbf{0.063}_{\pm0.013}$	$\mathbf{0.055}_{\pm0.007}$	$\mathbf{0.598}_{\pm0.028}$	$\mathbf{0.060}_{\pm0.007}$	$\mathbf{0.055}_{\pm0.008}$
		Far	$3.135_{\pm0.155}$	$\mathbf{0.102}_{\pm0.007}$	$\mathbf{0.033}_{\pm0.003}$	$3.209_{\pm0.135}$	$\mathbf{0.131}_{\pm0.010}$	$\mathbf{0.057}_{\pm0.008}$	$3.335_{\pm0.123}$	$\mathbf{0.115}_{\pm0.008}$	$\mathbf{0.056}_{\pm0.008}$

Table 13. **Ablation study of hyperparameter γ when generating pseudo OOD samples.** Estimation Error $(w-\hat{w})^2/K(\downarrow)$ of our OSLS estimation model (OpenMax OOD detector) on CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Forward) ID and OOD label shift.

Dataset			CIFAR100								
ID label Shift param			LT10 Backward			LT50 Backward			LT100 Backward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.540_{\pm0.029}$	$0.152_{\pm0.008}$	$0.159_{\pm0.025}$	$0.732_{\pm0.050}$	$0.252_{\pm0.027}$	$0.264_{\pm0.035}$	$0.778_{\pm0.026}$	$0.281_{\pm0.050}$	$0.339_{\pm0.073}$
		Far	$4.042_{\pm0.273}$	$0.276_{\pm0.011}$	$0.161_{\pm0.023}$	$4.075_{\pm0.388}$	$0.381_{\pm0.049}$	$0.262_{\pm0.037}$	$4.080_{\pm0.223}$	$0.387_{\pm0.056}$	$0.339_{\pm0.077}$
MLLS		Near	$0.912_{\pm0.083}$	$0.131_{\pm0.013}$	$0.119_{\pm0.009}$	$1.107_{\pm0.085}$	$0.203_{\pm0.012}$	$0.173_{\pm0.017}$	$1.152_{\pm0.061}$	$0.218_{\pm0.017}$	$0.203_{\pm0.022}$
		Far	$9.500_{\pm1.553}$	$0.332_{\pm0.036}$	$0.118_{\pm0.008}$	$9.583_{\pm1.578}$	$0.404_{\pm0.058}$	$0.167_{\pm0.017}$	$9.494_{\pm1.499}$	$0.381_{\pm0.039}$	$0.201_{\pm0.024}$
RLLS		Near	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.403_{\pm0.000}$	$1.402_{\pm0.000}$	$1.402_{\pm0.000}$
		Far	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.403_{\pm0.000}$	$1.402_{\pm0.000}$	$1.402_{\pm0.000}$
MAPLS		Near	$0.710_{\pm0.052}$	$0.119_{\pm0.007}$	$0.106_{\pm0.003}$	$0.941_{\pm0.063}$	$0.196_{\pm0.008}$	$0.159_{\pm0.009}$	$1.007_{\pm0.044}$	$0.218_{\pm0.012}$	$0.188_{\pm0.012}$
		Far	$7.360_{\pm1.206}$	$0.268_{\pm0.025}$	$0.106_{\pm0.002}$	$7.476_{\pm1.220}$	$0.345_{\pm0.039}$	$0.155_{\pm0.009}$	$7.439_{\pm1.153}$	$0.339_{\pm0.023}$	$0.186_{\pm0.012}$
Open Set Label Shift estimation models											
Baseline			$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$
ours	$\gamma=0.1$	Near	$0.428_{\pm0.046}$	$\mathbf{0.041}_{\pm0.002}$	$\mathbf{0.034}_{\pm0.004}$	$\mathbf{0.565}_{\pm0.076}$	$\mathbf{0.058}_{\pm0.001}$	$\mathbf{0.056}_{\pm0.006}$	$\mathbf{0.565}_{\pm0.028}$	$\mathbf{0.063}_{\pm0.001}$	$\mathbf{0.052}_{\pm0.003}$
		Far	$2.105_{\pm0.489}$	$\mathbf{0.080}_{\pm0.007}$	$\mathbf{0.035}_{\pm0.003}$	$2.244_{\pm0.372}$	$\mathbf{0.093}_{\pm0.004}$	$\mathbf{0.058}_{\pm0.007}$	$2.192_{\pm0.320}$	$\mathbf{0.100}_{\pm0.015}$	$\mathbf{0.054}_{\pm0.003}$
	$\gamma=0.2$	Near	$0.497_{\pm0.025}$	$\mathbf{0.034}_{\pm0.004}$	$\mathbf{0.028}_{\pm0.001}$	$\mathbf{0.628}_{\pm0.017}$	$\mathbf{0.059}_{\pm0.008}$	$\mathbf{0.049}_{\pm0.005}$	$\mathbf{0.664}_{\pm0.025}$	$\mathbf{0.058}_{\pm0.007}$	$\mathbf{0.046}_{\pm0.005}$
		Far	$2.879_{\pm0.178}$	$\mathbf{0.093}_{\pm0.007}$	$\mathbf{0.029}_{\pm0.001}$	$2.855_{\pm0.192}$	$\mathbf{0.115}_{\pm0.019}$	$\mathbf{0.050}_{\pm0.004}$	$2.879_{\pm0.215}$	$\mathbf{0.115}_{\pm0.010}$	$\mathbf{0.047}_{\pm0.005}$
	$\gamma=0.3$	Near	$0.539_{\pm0.012}$	$\mathbf{0.034}_{\pm0.004}$	$\mathbf{0.029}_{\pm0.006}$	$\mathbf{0.662}_{\pm0.014}$	$\mathbf{0.060}_{\pm0.010}$	$\mathbf{0.048}_{\pm0.003}$	$\mathbf{0.688}_{\pm0.021}$	$\mathbf{0.066}_{\pm0.007}$	$\mathbf{0.053}_{\pm0.003}$
		Far	$3.000_{\pm0.185}$	$\mathbf{0.092}_{\pm0.015}$	$\mathbf{0.030}_{\pm0.006}$	$2.987_{\pm0.171}$	$\mathbf{0.113}_{\pm0.006}$	$\mathbf{0.050}_{\pm0.003}$	$3.025_{\pm0.235}$	$\mathbf{0.114}_{\pm0.022}$	$\mathbf{0.055}_{\pm0.003}$
	$\gamma=0.4$	Near	$0.534_{\pm0.001}$	$\mathbf{0.037}_{\pm0.003}$	$\mathbf{0.027}_{\pm0.002}$	$\mathbf{0.669}_{\pm0.024}$	$\mathbf{0.059}_{\pm0.002}$	$\mathbf{0.049}_{\pm0.001}$	$\mathbf{0.696}_{\pm0.019}$	$\mathbf{0.063}_{\pm0.009}$	$\mathbf{0.046}_{\pm0.002}$
		Far	$3.001_{\pm0.144}$	$\mathbf{0.101}_{\pm0.009}$	$\mathbf{0.028}_{\pm0.003}$	$3.117_{\pm0.085}$	$\mathbf{0.113}_{\pm0.014}$	$\mathbf{0.050}_{\pm0.001}$	$3.023_{\pm0.134}$	$\mathbf{0.131}_{\pm0.016}$	$\mathbf{0.048}_{\pm0.003}$
	$\gamma=0.5$	Near	$0.526_{\pm0.018}$	$\mathbf{0.035}_{\pm0.002}$	$\mathbf{0.030}_{\pm0.004}$	$\mathbf{0.647}_{\pm0.005}$	$\mathbf{0.052}_{\pm0.004}$	$\mathbf{0.054}_{\pm0.005}$	$\mathbf{0.693}_{\pm0.017}$	$\mathbf{0.060}_{\pm0.002}$	$\mathbf{0.044}_{\pm0.004}$
		Far	$3.078_{\pm0.159}$	$\mathbf{0.092}_{\pm0.017}$	$\mathbf{0.032}_{\pm0.003}$	$3.046_{\pm0.153}$	$\mathbf{0.122}_{\pm0.018}$	$\mathbf{0.056}_{\pm0.006}$	$3.021_{\pm0.160}$	$\mathbf{0.118}_{\pm0.021}$	$\mathbf{0.046}_{\pm0.005}$

Table 14. **Ablation study of hyperparameter γ when generating pseudo OOD samples.** Estimation Error $(w-\hat{w})^2/K(\downarrow)$ of our OSLS estimation model (OpenMax OOD detector) on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Backward) ID and OOD label shift.

F. More Estimation Error Results

We provide the full open set label shift estimation error on the CIFAR10, CIFAR100 datasets. The results show that our model consistently performs better than the open set label shift baseline and CSLS estimation models on most of the tested dataset setups.

F.1. CIFAR10

Dataset			CIFAR10								
ID label Shift param			LT10 Forward			LT50 Forward			LT100 Forward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.172_{\pm 0.014}$	$0.005_{\pm 0.001}$	$0.003_{\pm 0.001}$	$0.341_{\pm 0.025}$	$0.008_{\pm 0.000}$	$0.007_{\pm 0.003}$	$0.418_{\pm 0.025}$	$0.011_{\pm 0.001}$	$0.010_{\pm 0.002}$
		Far	$0.400_{\pm 0.012}$	$0.014_{\pm 0.002}$	$0.004_{\pm 0.001}$	$0.547_{\pm 0.009}$	$0.016_{\pm 0.001}$	$0.008_{\pm 0.003}$	$0.629_{\pm 0.006}$	$0.018_{\pm 0.002}$	$0.010_{\pm 0.003}$
MLLS		Near	$0.168_{\pm 0.016}$	$0.005_{\pm 0.000}$	$0.002_{\pm 0.000}$	$0.318_{\pm 0.023}$	$0.007_{\pm 0.001}$	$0.002_{\pm 0.001}$	$0.388_{\pm 0.026}$	$0.010_{\pm 0.001}$	$0.002_{\pm 0.001}$
		Far	$0.458_{\pm 0.014}$	$0.012_{\pm 0.001}$	$0.002_{\pm 0.000}$	$0.587_{\pm 0.019}$	$0.013_{\pm 0.002}$	$0.002_{\pm 0.001}$	$0.662_{\pm 0.016}$	$0.015_{\pm 0.001}$	$0.002_{\pm 0.001}$
RLLS		Near	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$1.199_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.538_{\pm 0.000}$	$1.537_{\pm 0.000}$	$1.536_{\pm 0.000}$
		Far	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$1.199_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.538_{\pm 0.000}$	$1.537_{\pm 0.000}$	$1.536_{\pm 0.000}$
MAPLS		Near	$0.177_{\pm 0.014}$	$0.015_{\pm 0.001}$	$0.007_{\pm 0.001}$	$0.369_{\pm 0.022}$	$0.034_{\pm 0.003}$	$0.014_{\pm 0.002}$	$0.458_{\pm 0.024}$	$0.043_{\pm 0.004}$	$0.017_{\pm 0.003}$
		Far	$0.396_{\pm 0.013}$	$0.019_{\pm 0.001}$	$0.007_{\pm 0.001}$	$0.565_{\pm 0.019}$	$0.036_{\pm 0.003}$	$0.014_{\pm 0.002}$	$0.659_{\pm 0.016}$	$0.045_{\pm 0.003}$	$0.017_{\pm 0.003}$
Open Set Label Shift estimation models											
Baseline			$0.487_{\pm 0.000}$	$0.487_{\pm 0.000}$	$0.487_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.540_{\pm 0.000}$	$1.540_{\pm 0.000}$	$1.540_{\pm 0.000}$
ours	OpenMax	Near	$\mathbf{0.068}_{\pm 0.007}$	$\mathbf{0.002}_{\pm 0.001}$	$0.002_{\pm 0.000}$	$\mathbf{0.131}_{\pm 0.017}$	$\mathbf{0.003}_{\pm 0.001}$	$0.002_{\pm 0.000}$	$\mathbf{0.165}_{\pm 0.026}$	$\mathbf{0.003}_{\pm 0.000}$	$\mathbf{0.002}_{\pm 0.001}$
		Far	$\mathbf{0.186}_{\pm 0.005}$	$\mathbf{0.003}_{\pm 0.000}$	$0.002_{\pm 0.000}$	$\mathbf{0.244}_{\pm 0.006}$	$\mathbf{0.005}_{\pm 0.001}$	$\mathbf{0.002}_{\pm 0.000}$	$\mathbf{0.264}_{\pm 0.009}$	$\mathbf{0.004}_{\pm 0.000}$	$\mathbf{0.002}_{\pm 0.000}$
	MLS	Near	$\mathbf{0.026}_{\pm 0.001}$	$0.007_{\pm 0.001}$	$0.007_{\pm 0.001}$	$\mathbf{0.056}_{\pm 0.004}$	$0.007_{\pm 0.001}$	$0.010_{\pm 0.001}$	$\mathbf{0.067}_{\pm 0.008}$	$\mathbf{0.009}_{\pm 0.002}$	$0.008_{\pm 0.002}$
		Far	$\mathbf{0.042}_{\pm 0.012}$	$\mathbf{0.007}_{\pm 0.001}$	$0.007_{\pm 0.001}$	$\mathbf{0.057}_{\pm 0.012}$	$\mathbf{0.008}_{\pm 0.000}$	$0.010_{\pm 0.001}$	$\mathbf{0.061}_{\pm 0.013}$	$\mathbf{0.009}_{\pm 0.002}$	$0.008_{\pm 0.002}$
	ReAct	Near	$\mathbf{0.060}_{\pm 0.015}$	$0.017_{\pm 0.010}$	$0.017_{\pm 0.012}$	$\mathbf{0.098}_{\pm 0.000}$	$0.026_{\pm 0.012}$	$0.023_{\pm 0.011}$	$\mathbf{0.114}_{\pm 0.005}$	$0.029_{\pm 0.020}$	$0.028_{\pm 0.017}$
		Far	$\mathbf{0.084}_{\pm 0.034}$	$0.017_{\pm 0.010}$	$0.017_{\pm 0.012}$	$\mathbf{0.110}_{\pm 0.033}$	$0.026_{\pm 0.012}$	$0.023_{\pm 0.011}$	$\mathbf{0.119}_{\pm 0.037}$	$0.029_{\pm 0.021}$	$0.028_{\pm 0.017}$
	KNN	Near	$\mathbf{0.035}_{\pm 0.003}$	$0.006_{\pm 0.002}$	$0.006_{\pm 0.002}$	$\mathbf{0.071}_{\pm 0.005}$	$\mathbf{0.005}_{\pm 0.001}$	$0.008_{\pm 0.001}$	$\mathbf{0.090}_{\pm 0.011}$	$\mathbf{0.006}_{\pm 0.001}$	$0.005_{\pm 0.001}$
		Far	$\mathbf{0.075}_{\pm 0.016}$	$\mathbf{0.006}_{\pm 0.002}$	$0.006_{\pm 0.002}$	$\mathbf{0.090}_{\pm 0.010}$	$\mathbf{0.006}_{\pm 0.001}$	$0.008_{\pm 0.002}$	$\mathbf{0.104}_{\pm 0.015}$	$\mathbf{0.006}_{\pm 0.001}$	$0.005_{\pm 0.001}$
	Ash	Near	$0.280_{\pm 0.086}$	$0.221_{\pm 0.124}$	$0.211_{\pm 0.128}$	$0.412_{\pm 0.127}$	$0.319_{\pm 0.209}$	$0.332_{\pm 0.217}$	$0.452_{\pm 0.135}$	$0.359_{\pm 0.224}$	$0.366_{\pm 0.241}$
		Far	$\mathbf{0.393}_{\pm 0.132}$	$0.225_{\pm 0.126}$	$0.211_{\pm 0.128}$	$\mathbf{0.521}_{\pm 0.172}$	$0.322_{\pm 0.211}$	$0.333_{\pm 0.217}$	$\mathbf{0.568}_{\pm 0.181}$	$0.364_{\pm 0.228}$	$0.367_{\pm 0.242}$

Table 15. **Estimation Error $(w - \hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR10 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Forward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among corresponding the OOD test set (Tab. 1) and over three independent ID classifiers.

Dataset			CIFAR10								
ID label Shift param			LT10 Backward			LT50 Backward			LT100 Backward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.232_{\pm 0.023}$	$0.008_{\pm 0.003}$	$0.001_{\pm 0.000}$	$0.433_{\pm 0.034}$	$0.017_{\pm 0.002}$	$0.002_{\pm 0.001}$	$0.508_{\pm 0.029}$	$0.021_{\pm 0.002}$	$0.002_{\pm 0.001}$
		Far	$0.568_{\pm 0.032}$	$0.020_{\pm 0.003}$	$0.001_{\pm 0.000}$	$0.796_{\pm 0.030}$	$0.029_{\pm 0.003}$	$0.002_{\pm 0.001}$	$0.905_{\pm 0.037}$	$0.034_{\pm 0.002}$	$0.002_{\pm 0.001}$
MLLS		Near	$0.222_{\pm 0.030}$	$0.009_{\pm 0.004}$	$0.002_{\pm 0.001}$	$0.394_{\pm 0.049}$	$0.016_{\pm 0.005}$	$0.003_{\pm 0.001}$	$0.455_{\pm 0.045}$	$0.019_{\pm 0.006}$	$0.003_{\pm 0.002}$
		Far	$0.644_{\pm 0.017}$	$0.021_{\pm 0.004}$	$0.002_{\pm 0.001}$	$0.857_{\pm 0.020}$	$0.030_{\pm 0.005}$	$0.003_{\pm 0.002}$	$0.968_{\pm 0.024}$	$0.034_{\pm 0.006}$	$0.004_{\pm 0.002}$
RLLS		Near	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$1.199_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.538_{\pm 0.000}$	$1.537_{\pm 0.000}$	$1.537_{\pm 0.000}$
		Far	$0.487_{\pm 0.000}$	$0.486_{\pm 0.000}$	$0.486_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.198_{\pm 0.000}$	$1.538_{\pm 0.000}$	$1.537_{\pm 0.000}$	$1.537_{\pm 0.000}$
MAPLS		Near	$0.233_{\pm 0.027}$	$0.023_{\pm 0.005}$	$0.011_{\pm 0.002}$	$0.447_{\pm 0.046}$	$0.052_{\pm 0.008}$	$0.025_{\pm 0.004}$	$0.530_{\pm 0.042}$	$0.063_{\pm 0.010}$	$0.032_{\pm 0.006}$
		Far	$0.574_{\pm 0.018}$	$0.035_{\pm 0.005}$	$0.011_{\pm 0.002}$	$0.824_{\pm 0.023}$	$0.066_{\pm 0.008}$	$0.025_{\pm 0.005}$	$0.949_{\pm 0.028}$	$0.080_{\pm 0.010}$	$0.032_{\pm 0.006}$
Open Set Label Shift estimation models											
Baseline			$0.487_{\pm 0.000}$	$0.487_{\pm 0.000}$	$0.487_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.200_{\pm 0.000}$	$1.540_{\pm 0.000}$	$1.540_{\pm 0.000}$	$1.540_{\pm 0.000}$
ours	OpenMax	Near	$\mathbf{0.090}_{\pm 0.013}$	$\mathbf{0.002}_{\pm 0.000}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.158}_{\pm 0.022}$	$\mathbf{0.004}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.186}_{\pm 0.017}$	$\mathbf{0.004}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.000}$
		Far	$\mathbf{0.225}_{\pm 0.009}$	$\mathbf{0.004}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.304}_{\pm 0.023}$	$\mathbf{0.006}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.326}_{\pm 0.022}$	$\mathbf{0.007}_{\pm 0.002}$	$\mathbf{0.001}_{\pm 0.000}$
	MLS	Near	$\mathbf{0.021}_{\pm 0.004}$	$\mathbf{0.002}_{\pm 0.001}$	$0.003_{\pm 0.001}$	$\mathbf{0.045}_{\pm 0.013}$	$\mathbf{0.002}_{\pm 0.001}$	$0.002_{\pm 0.001}$	$\mathbf{0.058}_{\pm 0.010}$	$\mathbf{0.002}_{\pm 0.000}$	$\mathbf{0.002}_{\pm 0.001}$
		Far	$\mathbf{0.035}_{\pm 0.011}$	$\mathbf{0.003}_{\pm 0.001}$	$0.003_{\pm 0.001}$	$\mathbf{0.044}_{\pm 0.014}$	$\mathbf{0.002}_{\pm 0.001}$	$0.002_{\pm 0.001}$	$\mathbf{0.052}_{\pm 0.017}$	$\mathbf{0.002}_{\pm 0.000}$	$\mathbf{0.002}_{\pm 0.001}$
	ReAct	Near	$\mathbf{0.048}_{\pm 0.016}$	$\mathbf{0.004}_{\pm 0.001}$	$0.003_{\pm 0.001}$	$\mathbf{0.086}_{\pm 0.034}$	$\mathbf{0.005}_{\pm 0.004}$	$0.004_{\pm 0.003}$	$\mathbf{0.101}_{\pm 0.036}$	$\mathbf{0.005}_{\pm 0.004}$	$0.003_{\pm 0.003}$
		Far	$\mathbf{0.073}_{\pm 0.044}$	$\mathbf{0.005}_{\pm 0.001}$	$0.003_{\pm 0.001}$	$\mathbf{0.093}_{\pm 0.050}$	$\mathbf{0.005}_{\pm 0.004}$	$0.004_{\pm 0.003}$	$\mathbf{0.103}_{\pm 0.057}$	$\mathbf{0.005}_{\pm 0.004}$	$0.003_{\pm 0.003}$
	KNN	Near	$\mathbf{0.033}_{\pm 0.012}$	$\mathbf{0.001}_{\pm 0.000}$	$0.002_{\pm 0.000}$	$\mathbf{0.066}_{\pm 0.018}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.086}_{\pm 0.026}$	$\mathbf{0.002}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.001}$
		Far	$\mathbf{0.075}_{\pm 0.021}$	$\mathbf{0.002}_{\pm 0.001}$	$0.002_{\pm 0.000}$	$\mathbf{0.100}_{\pm 0.025}$	$\mathbf{0.001}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.000}$	$\mathbf{0.101}_{\pm 0.031}$	$\mathbf{0.002}_{\pm 0.001}$	$\mathbf{0.001}_{\pm 0.001}$
	Ash	Near	$0.447_{\pm 0.433}$	$0.374_{\pm 0.443}$	$0.380_{\pm 0.441}$	$0.797_{\pm 0.836}$	$0.693_{\pm 0.875}$	$0.705_{\pm 0.917}$	$0.969_{\pm 1.051}$	$0.891_{\pm 1.157}$	$0.901_{\pm 1.173}$
		Far	$0.597_{\pm 0.442}$	$0.378_{\pm 0.444}$	$0.380_{\pm 0.440}$	$0.960_{\pm 0.869}$	$0.698_{\pm 0.878}$	$0.705_{\pm 0.916}$	$1.168_{\pm 1.128}$	$0.903_{\pm 1.164}$	$0.903_{\pm 1.175}$

Table 16. **Estimation Error $(w-\hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR10 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Backward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

Dataset			CIFAR10					
ID label Shift param			Dir $\alpha = 1.0$			Dir $\alpha = 10.0$		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models								
BBSE		Near	$0.309_{\pm 0.103}$	$0.008_{\pm 0.002}$	$0.001_{\pm 0.000}$	$0.079_{\pm 0.039}$	$0.003_{\pm 0.001}$	$0.002_{\pm 0.001}$
		Far	$0.570_{\pm 0.114}$	$0.016_{\pm 0.005}$	$0.001_{\pm 0.000}$	$0.366_{\pm 0.087}$	$0.012_{\pm 0.001}$	$0.002_{\pm 0.001}$
MLLS		Near	$0.280_{\pm 0.082}$	$0.010_{\pm 0.004}$	$0.002_{\pm 0.001}$	$0.088_{\pm 0.040}$	$0.003_{\pm 0.001}$	$0.001_{\pm 0.000}$
		Far	$0.616_{\pm 0.073}$	$0.016_{\pm 0.005}$	$0.002_{\pm 0.001}$	$0.439_{\pm 0.074}$	$0.012_{\pm 0.000}$	$0.001_{\pm 0.000}$
RLLS		Near	$0.979_{\pm 0.597}$	$0.800_{\pm 0.070}$	$0.738_{\pm 0.248}$	$0.082_{\pm 0.020}$	$0.074_{\pm 0.023}$	$0.075_{\pm 0.030}$
		Far	$0.979_{\pm 0.597}$	$0.800_{\pm 0.070}$	$0.738_{\pm 0.248}$	$0.082_{\pm 0.020}$	$0.074_{\pm 0.023}$	$0.075_{\pm 0.030}$
MAPLS		Near	$0.319_{\pm 0.115}$	$0.033_{\pm 0.009}$	$0.014_{\pm 0.005}$	$0.074_{\pm 0.037}$	$0.003_{\pm 0.001}$	$0.002_{\pm 0.001}$
		Far	$0.577_{\pm 0.110}$	$0.036_{\pm 0.010}$	$0.014_{\pm 0.005}$	$0.351_{\pm 0.071}$	$0.010_{\pm 0.001}$	$0.002_{\pm 0.001}$
Open Set Label Shift estimation models								
Baseline			$0.980_{\pm 0.598}$	$0.801_{\pm 0.070}$	$0.740_{\pm 0.249}$	$0.082_{\pm 0.020}$	$0.074_{\pm 0.023}$	$0.075_{\pm 0.030}$
ours	OpenMax	Near	$\mathbf{0.066}_{\pm 0.029}$	$0.006_{\pm 0.003}$	$0.004_{\pm 0.005}$	$\mathbf{0.030}_{\pm 0.009}$	$\mathbf{0.002}_{\pm 0.001}$	$0.002_{\pm 0.000}$
		Far	$\mathbf{0.174}_{\pm 0.055}$	$\mathbf{0.006}_{\pm 0.001}$	$0.004_{\pm 0.005}$	$0.130_{\pm 0.033}$	$\mathbf{0.003}_{\pm 0.000}$	$0.002_{\pm 0.000}$
	MLS	Near	$\mathbf{0.041}_{\pm 0.008}$	$0.010_{\pm 0.004}$	$0.006_{\pm 0.000}$	$\mathbf{0.010}_{\pm 0.003}$	$0.004_{\pm 0.002}$	$0.005_{\pm 0.001}$
		Far	$\mathbf{0.055}_{\pm 0.009}$	$\mathbf{0.009}_{\pm 0.004}$	$0.006_{\pm 0.000}$	$\mathbf{0.034}_{\pm 0.013}$	$\mathbf{0.004}_{\pm 0.002}$	$0.005_{\pm 0.001}$
	ReAct	Near	$\mathbf{0.092}_{\pm 0.031}$	$0.014_{\pm 0.013}$	$0.004_{\pm 0.001}$	$\mathbf{0.032}_{\pm 0.018}$	$0.004_{\pm 0.002}$	$0.004_{\pm 0.001}$
		Far	$\mathbf{0.108}_{\pm 0.051}$	$\mathbf{0.015}_{\pm 0.013}$	$0.004_{\pm 0.001}$	$\mathbf{0.062}_{\pm 0.039}$	$\mathbf{0.005}_{\pm 0.002}$	$0.004_{\pm 0.001}$
	KNN	Near	$\mathbf{0.029}_{\pm 0.009}$	$\mathbf{0.004}_{\pm 0.003}$	$0.012_{\pm 0.006}$	$\mathbf{0.009}_{\pm 0.002}$	$0.004_{\pm 0.001}$	$0.008_{\pm 0.003}$
		Far	$\mathbf{0.051}_{\pm 0.012}$	$\mathbf{0.004}_{\pm 0.003}$	$0.012_{\pm 0.006}$	$\mathbf{0.034}_{\pm 0.004}$	$\mathbf{0.005}_{\pm 0.002}$	$0.008_{\pm 0.003}$
	Ash	Near	$0.284_{\pm 0.174}$	$0.103_{\pm 0.072}$	$0.619_{\pm 0.395}$	$0.174_{\pm 0.098}$	$0.118_{\pm 0.040}$	$0.188_{\pm 0.060}$
		Far	$\mathbf{0.405}_{\pm 0.154}$	$0.107_{\pm 0.075}$	$0.619_{\pm 0.394}$	$0.307_{\pm 0.117}$	$0.123_{\pm 0.041}$	$0.188_{\pm 0.060}$

Table 17. **Estimation Error $(w-\hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR10 dataset with Near OOD datasets and Far OOD datasets comparison under Dirichlet ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

F.2. CIFAR100

Dataset			CIFAR100								
ID label Shift param			LT10 Forward			LT50 Forward			LT100 Forward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.529_{\pm0.017}$	$0.131_{\pm0.030}$	$0.097_{\pm0.021}$	$0.750_{\pm0.033}$	$0.173_{\pm0.010}$	$0.176_{\pm0.058}$	$0.850_{\pm0.052}$	$0.173_{\pm0.041}$	$0.167_{\pm0.054}$
		Far	$4.118_{\pm0.263}$	$0.250_{\pm0.037}$	$0.099_{\pm0.021}$	$4.362_{\pm0.245}$	$0.290_{\pm0.017}$	$0.176_{\pm0.057}$	$4.489_{\pm0.238}$	$0.294_{\pm0.039}$	$0.168_{\pm0.053}$
MLLS		Near	$0.870_{\pm0.069}$	$0.116_{\pm0.019}$	$0.080_{\pm0.022}$	$1.005_{\pm0.085}$	$0.142_{\pm0.017}$	$0.116_{\pm0.030}$	$1.100_{\pm0.098}$	$0.132_{\pm0.034}$	$0.113_{\pm0.039}$
		Far	$9.656_{\pm1.747}$	$0.328_{\pm0.034}$	$0.083_{\pm0.023}$	$9.804_{\pm1.365}$	$0.364_{\pm0.044}$	$0.119_{\pm0.029}$	$9.862_{\pm1.469}$	$0.353_{\pm0.044}$	$0.117_{\pm0.040}$
RLLS		Near	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.404_{\pm0.000}$	$1.402_{\pm0.000}$	$1.402_{\pm0.000}$
		Far	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.404_{\pm0.000}$	$1.403_{\pm0.000}$	$1.402_{\pm0.000}$
MAPLS		Near	$0.672_{\pm0.040}$	$0.116_{\pm0.014}$	$0.085_{\pm0.014}$	$0.860_{\pm0.058}$	$0.161_{\pm0.013}$	$0.129_{\pm0.019}$	$0.965_{\pm0.060}$	$0.164_{\pm0.024}$	$0.134_{\pm0.026}$
		Far	$7.481_{\pm1.351}$	$0.275_{\pm0.024}$	$0.087_{\pm0.014}$	$7.667_{\pm1.050}$	$0.330_{\pm0.031}$	$0.131_{\pm0.018}$	$7.763_{\pm1.137}$	$0.336_{\pm0.027}$	$0.137_{\pm0.027}$
Open Set Label Shift estimation models											
Baseline			$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$
ours	OpenMax	Near	$\mathbf{0.387}_{\pm0.030}$	$\mathbf{0.043}_{\pm0.006}$	$\mathbf{0.046}_{\pm0.009}$	$\mathbf{0.450}_{\pm0.033}$	$\mathbf{0.071}_{\pm0.008}$	$\mathbf{0.071}_{\pm0.017}$	$\mathbf{0.511}_{\pm0.038}$	$\mathbf{0.081}_{\pm0.012}$	$\mathbf{0.077}_{\pm0.005}$
		Far	$2.223_{\pm0.233}$	$\mathbf{0.087}_{\pm0.002}$	$\mathbf{0.046}_{\pm0.010}$	$2.353_{\pm0.164}$	$\mathbf{0.111}_{\pm0.008}$	$\mathbf{0.073}_{\pm0.017}$	$2.341_{\pm0.498}$	$\mathbf{0.118}_{\pm0.015}$	$\mathbf{0.078}_{\pm0.005}$
	MLS	Near	$\mathbf{0.323}_{\pm0.020}$	$\mathbf{0.078}_{\pm0.004}$	$\mathbf{0.078}_{\pm0.005}$	$\mathbf{0.371}_{\pm0.011}$	$\mathbf{0.120}_{\pm0.011}$	$0.121_{\pm0.011}$	$\mathbf{0.415}_{\pm0.017}$	$\mathbf{0.126}_{\pm0.014}$	$\mathbf{0.120}_{\pm0.013}$
		Far	$1.289_{\pm0.337}$	$\mathbf{0.099}_{\pm0.008}$	$\mathbf{0.079}_{\pm0.006}$	$1.324_{\pm0.276}$	$\mathbf{0.138}_{\pm0.015}$	$0.123_{\pm0.011}$	$\mathbf{1.366}_{\pm0.313}$	$\mathbf{0.150}_{\pm0.015}$	$\mathbf{0.120}_{\pm0.013}$
	ReAct	Near	$\mathbf{0.331}_{\pm0.030}$	$\mathbf{0.076}_{\pm0.006}$	$\mathbf{0.075}_{\pm0.005}$	$\mathbf{0.362}_{\pm0.016}$	$\mathbf{0.114}_{\pm0.008}$	$0.116_{\pm0.021}$	$\mathbf{0.396}_{\pm0.023}$	$\mathbf{0.110}_{\pm0.012}$	$0.124_{\pm0.010}$
		Far	$1.138_{\pm0.251}$	$\mathbf{0.096}_{\pm0.004}$	$\mathbf{0.075}_{\pm0.005}$	$1.199_{\pm0.302}$	$\mathbf{0.131}_{\pm0.008}$	$0.117_{\pm0.022}$	$\mathbf{1.202}_{\pm0.281}$	$\mathbf{0.127}_{\pm0.008}$	$0.124_{\pm0.010}$
	KNN	Near	$0.736_{\pm0.026}$	$0.141_{\pm0.008}$	$0.139_{\pm0.002}$	$0.805_{\pm0.023}$	$0.206_{\pm0.031}$	$0.205_{\pm0.030}$	$\mathbf{0.817}_{\pm0.034}$	$0.235_{\pm0.018}$	$0.228_{\pm0.023}$
		Far	$1.188_{\pm0.173}$	$\mathbf{0.152}_{\pm0.006}$	$0.140_{\pm0.002}$	$1.281_{\pm0.161}$	$\mathbf{0.216}_{\pm0.033}$	$0.207_{\pm0.030}$	$\mathbf{1.287}_{\pm0.126}$	$\mathbf{0.246}_{\pm0.016}$	$0.229_{\pm0.023}$
	Ash	Near	$\mathbf{0.358}_{\pm0.050}$	$\mathbf{0.111}_{\pm0.005}$	$0.101_{\pm0.015}$	$\mathbf{0.514}_{\pm0.028}$	$0.194_{\pm0.037}$	$0.176_{\pm0.027}$	$\mathbf{0.541}_{\pm0.051}$	$0.217_{\pm0.023}$	$0.198_{\pm0.035}$
		Far	$1.015_{\pm0.121}$	$\mathbf{0.119}_{\pm0.002}$	$0.101_{\pm0.014}$	$1.169_{\pm0.245}$	$\mathbf{0.206}_{\pm0.031}$	$0.175_{\pm0.027}$	$\mathbf{1.183}_{\pm0.120}$	$\mathbf{0.229}_{\pm0.021}$	$0.198_{\pm0.035}$

Table 18. **Estimation Error $(w - \hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Forward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

Dataset			CIFAR100								
ID label Shift param			LT10 Backward			LT50 Backward			LT100 Backward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.553_{\pm0.020}$	$0.153_{\pm0.003}$	$0.162_{\pm0.016}$	$0.730_{\pm0.049}$	$0.243_{\pm0.031}$	$0.244_{\pm0.022}$	$0.778_{\pm0.043}$	$0.265_{\pm0.017}$	$0.295_{\pm0.039}$
		Far	$3.968_{\pm0.256}$	$0.276_{\pm0.014}$	$0.164_{\pm0.016}$	$4.099_{\pm0.290}$	$0.351_{\pm0.030}$	$0.245_{\pm0.024}$	$4.104_{\pm0.206}$	$0.373_{\pm0.028}$	$0.296_{\pm0.042}$
MLLS		Near	$0.937_{\pm0.034}$	$0.140_{\pm0.016}$	$0.116_{\pm0.009}$	$1.133_{\pm0.107}$	$0.208_{\pm0.021}$	$0.171_{\pm0.015}$	$1.146_{\pm0.083}$	$0.212_{\pm0.038}$	$0.193_{\pm0.011}$
		Far	$9.372_{\pm1.494}$	$0.338_{\pm0.043}$	$0.116_{\pm0.008}$	$9.548_{\pm1.680}$	$0.380_{\pm0.042}$	$0.168_{\pm0.011}$	$9.463_{\pm1.382}$	$0.377_{\pm0.015}$	$0.188_{\pm0.010}$
RLLS		Near	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.403_{\pm0.000}$	$1.402_{\pm0.000}$	$1.402_{\pm0.000}$
		Far	$0.426_{\pm0.000}$	$0.425_{\pm0.000}$	$0.425_{\pm0.000}$	$1.100_{\pm0.000}$	$1.099_{\pm0.000}$	$1.099_{\pm0.000}$	$1.403_{\pm0.000}$	$1.402_{\pm0.000}$	$1.402_{\pm0.000}$
MAPLS		Near	$0.729_{\pm0.019}$	$0.126_{\pm0.012}$	$0.103_{\pm0.007}$	$0.955_{\pm0.078}$	$0.199_{\pm0.013}$	$0.157_{\pm0.011}$	$1.005_{\pm0.063}$	$0.212_{\pm0.028}$	$0.185_{\pm0.006}$
		Far	$7.263_{\pm1.162}$	$0.273_{\pm0.029}$	$0.103_{\pm0.007}$	$7.452_{\pm1.294}$	$0.326_{\pm0.027}$	$0.155_{\pm0.009}$	$7.417_{\pm1.065}$	$0.334_{\pm0.006}$	$0.181_{\pm0.005}$
Open Set Label Shift estimation models											
Baseline			$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$0.426_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.101_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$	$1.405_{\pm0.000}$
ours	OpenMax	Near	$0.435_{\pm0.009}$	$\mathbf{0.039}_{\pm0.002}$	$\mathbf{0.031}_{\pm0.002}$	$\mathbf{0.528}_{\pm0.052}$	$\mathbf{0.061}_{\pm0.002}$	$\mathbf{0.051}_{\pm0.003}$	$\mathbf{0.607}_{\pm0.050}$	$\mathbf{0.068}_{\pm0.007}$	$\mathbf{0.058}_{\pm0.002}$
		Far	$2.174_{\pm0.378}$	$\mathbf{0.078}_{\pm0.012}$	$\mathbf{0.032}_{\pm0.003}$	$2.127_{\pm0.242}$	$\mathbf{0.105}_{\pm0.031}$	$\mathbf{0.051}_{\pm0.003}$	$2.175_{\pm0.336}$	$\mathbf{0.099}_{\pm0.012}$	$\mathbf{0.059}_{\pm0.002}$
	MLS	Near	$\mathbf{0.350}_{\pm0.034}$	$\mathbf{0.073}_{\pm0.001}$	$\mathbf{0.066}_{\pm0.004}$	$\mathbf{0.422}_{\pm0.041}$	$\mathbf{0.094}_{\pm0.009}$	$\mathbf{0.091}_{\pm0.001}$	$\mathbf{0.397}_{\pm0.026}$	$\mathbf{0.103}_{\pm0.003}$	$\mathbf{0.095}_{\pm0.010}$
		Far	$1.214_{\pm0.298}$	$\mathbf{0.087}_{\pm0.003}$	$\mathbf{0.067}_{\pm0.004}$	$1.169_{\pm0.301}$	$\mathbf{0.109}_{\pm0.006}$	$\mathbf{0.091}_{\pm0.002}$	$\mathbf{1.257}_{\pm0.287}$	$\mathbf{0.118}_{\pm0.010}$	$\mathbf{0.096}_{\pm0.011}$
	ReAct	Near	$\mathbf{0.324}_{\pm0.022}$	$\mathbf{0.066}_{\pm0.003}$	$\mathbf{0.070}_{\pm0.008}$	$\mathbf{0.415}_{\pm0.044}$	$\mathbf{0.092}_{\pm0.006}$	$\mathbf{0.084}_{\pm0.002}$	$\mathbf{0.402}_{\pm0.008}$	$\mathbf{0.100}_{\pm0.006}$	$\mathbf{0.087}_{\pm0.005}$
		Far	$1.009_{\pm0.211}$	$\mathbf{0.077}_{\pm0.003}$	$\mathbf{0.070}_{\pm0.008}$	$\mathbf{1.016}_{\pm0.263}$	$\mathbf{0.103}_{\pm0.002}$	$\mathbf{0.084}_{\pm0.003}$	$\mathbf{1.013}_{\pm0.236}$	$\mathbf{0.108}_{\pm0.007}$	$\mathbf{0.087}_{\pm0.005}$
	KNN	Near	$0.685_{\pm0.053}$	$0.131_{\pm0.002}$	$0.122_{\pm0.003}$	$0.740_{\pm0.011}$	$\mathbf{0.166}_{\pm0.007}$	$\mathbf{0.142}_{\pm0.012}$	$0.806_{\pm0.047}$	$\mathbf{0.160}_{\pm0.002}$	$\mathbf{0.145}_{\pm0.003}$
		Far	$1.091_{\pm0.141}$	$\mathbf{0.136}_{\pm0.002}$	$0.122_{\pm0.003}$	$\mathbf{1.099}_{\pm0.159}$	$\mathbf{0.171}_{\pm0.008}$	$\mathbf{0.141}_{\pm0.012}$	$\mathbf{1.112}_{\pm0.150}$	$\mathbf{0.161}_{\pm0.003}$	$\mathbf{0.145}_{\pm0.002}$
	Ash	Near	$\mathbf{0.388}_{\pm0.037}$	$\mathbf{0.079}_{\pm0.008}$	$\mathbf{0.068}_{\pm0.008}$	$\mathbf{0.482}_{\pm0.047}$	$\mathbf{0.112}_{\pm0.008}$	$\mathbf{0.098}_{\pm0.007}$	$\mathbf{0.514}_{\pm0.036}$	$\mathbf{0.115}_{\pm0.016}$	$\mathbf{0.121}_{\pm0.021}$
		Far	$0.933_{\pm0.090}$	$\mathbf{0.090}_{\pm0.007}$	$\mathbf{0.068}_{\pm0.008}$	$\mathbf{0.954}_{\pm0.089}$	$\mathbf{0.121}_{\pm0.007}$	$\mathbf{0.098}_{\pm0.007}$	$\mathbf{0.992}_{\pm0.122}$	$\mathbf{0.118}_{\pm0.014}$	$\mathbf{0.122}_{\pm0.021}$

Table 19. **Estimation Error $(w - \hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Backward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

Dataset			CIFAR100					
ID label Shift param			Dir $\alpha = 1.0$			Dir $\alpha = 10.0$		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models								
BBSE		Near	$0.659_{\pm 0.011}$	$0.138_{\pm 0.018}$	$0.207_{\pm 0.017}$	$0.466_{\pm 0.051}$	$0.108_{\pm 0.009}$	$0.106_{\pm 0.013}$
		Far	$4.243_{\pm 0.232}$	$0.263_{\pm 0.011}$	$0.210_{\pm 0.019}$	$3.874_{\pm 0.157}$	$0.240_{\pm 0.015}$	$0.110_{\pm 0.013}$
MLLS		Near	$1.006_{\pm 0.072}$	$0.139_{\pm 0.002}$	$0.162_{\pm 0.047}$	$0.845_{\pm 0.108}$	$0.112_{\pm 0.012}$	$0.095_{\pm 0.022}$
		Far	$9.871_{\pm 1.504}$	$0.369_{\pm 0.030}$	$0.162_{\pm 0.046}$	$9.281_{\pm 1.478}$	$0.337_{\pm 0.047}$	$0.097_{\pm 0.021}$
RLLS		Near	$1.087_{\pm 0.269}$	$0.888_{\pm 0.126}$	$1.119_{\pm 0.196}$	$0.105_{\pm 0.008}$	$0.105_{\pm 0.016}$	$0.096_{\pm 0.016}$
		Far	$1.088_{\pm 0.269}$	$0.888_{\pm 0.126}$	$1.119_{\pm 0.196}$	$0.105_{\pm 0.008}$	$0.105_{\pm 0.016}$	$0.096_{\pm 0.016}$
MAPLS		Near	$0.845_{\pm 0.055}$	$0.154_{\pm 0.016}$	$0.190_{\pm 0.057}$	$0.618_{\pm 0.073}$	$0.098_{\pm 0.008}$	$0.084_{\pm 0.008}$
		Far	$7.718_{\pm 1.108}$	$0.326_{\pm 0.036}$	$0.190_{\pm 0.056}$	$7.152_{\pm 1.138}$	$0.261_{\pm 0.034}$	$0.085_{\pm 0.008}$
Open Set Label Shift estimation models								
Baseline			$1.089_{\pm 0.270}$	$0.890_{\pm 0.126}$	$1.121_{\pm 0.196}$	$0.105_{\pm 0.008}$	$0.105_{\pm 0.016}$	$0.096_{\pm 0.016}$
ours	OpenMax	Near	$\mathbf{0.479}_{\pm 0.016}$	$\mathbf{0.049}_{\pm 0.008}$	$\mathbf{0.062}_{\pm 0.029}$	$0.365_{\pm 0.030}$	$\mathbf{0.042}_{\pm 0.003}$	$\mathbf{0.040}_{\pm 0.004}$
		Far	$2.282_{\pm 0.352}$	$\mathbf{0.087}_{\pm 0.003}$	$\mathbf{0.063}_{\pm 0.029}$	$2.034_{\pm 0.423}$	$\mathbf{0.075}_{\pm 0.007}$	$\mathbf{0.040}_{\pm 0.003}$
	MLS	Near	$\mathbf{0.358}_{\pm 0.028}$	$\mathbf{0.073}_{\pm 0.005}$	$0.145_{\pm 0.036}$	$0.290_{\pm 0.007}$	$\mathbf{0.077}_{\pm 0.005}$	$\mathbf{0.069}_{\pm 0.002}$
		Far	$1.232_{\pm 0.279}$	$\mathbf{0.087}_{\pm 0.009}$	$0.146_{\pm 0.035}$	$1.225_{\pm 0.315}$	$\mathbf{0.098}_{\pm 0.006}$	$\mathbf{0.069}_{\pm 0.002}$
	ReAct	Near	$\mathbf{0.379}_{\pm 0.024}$	$\mathbf{0.108}_{\pm 0.028}$	$\mathbf{0.080}_{\pm 0.031}$	$0.319_{\pm 0.026}$	$\mathbf{0.071}_{\pm 0.003}$	$\mathbf{0.074}_{\pm 0.006}$
		Far	$1.111_{\pm 0.260}$	$\mathbf{0.119}_{\pm 0.028}$	$\mathbf{0.081}_{\pm 0.031}$	$0.966_{\pm 0.268}$	$\mathbf{0.086}_{\pm 0.003}$	$\mathbf{0.074}_{\pm 0.005}$
	KNN	Near	$0.884_{\pm 0.114}$	$0.175_{\pm 0.047}$	$0.195_{\pm 0.027}$	$0.695_{\pm 0.034}$	$0.148_{\pm 0.024}$	$0.128_{\pm 0.017}$
		Far	$1.348_{\pm 0.252}$	$\mathbf{0.185}_{\pm 0.046}$	$0.195_{\pm 0.027}$	$1.103_{\pm 0.120}$	$0.156_{\pm 0.020}$	$0.128_{\pm 0.018}$
	Ash	Near	$\mathbf{0.406}_{\pm 0.052}$	$\mathbf{0.091}_{\pm 0.016}$	$\mathbf{0.096}_{\pm 0.010}$	$0.370_{\pm 0.031}$	$\mathbf{0.077}_{\pm 0.002}$	$\mathbf{0.074}_{\pm 0.011}$
		Far	$\mathbf{1.025}_{\pm 0.232}$	$\mathbf{0.098}_{\pm 0.021}$	$\mathbf{0.096}_{\pm 0.010}$	$0.935_{\pm 0.124}$	$\mathbf{0.090}_{\pm 0.006}$	$\mathbf{0.075}_{\pm 0.011}$

Table 20. **Estimation Error $(w - \hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Dirichlet ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

F.3. ImageNet-200

Dataset			ImageNet-200								
ID label Shift param			LT10 Forward			LT50 Forward			LT100 Forward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.564_{\pm 0.014}$	$0.119_{\pm 0.015}$	$0.107_{\pm 0.012}$	$0.664_{\pm 0.030}$	$0.128_{\pm 0.016}$	$0.115_{\pm 0.012}$	$0.735_{\pm 0.040}$	$0.132_{\pm 0.009}$	$0.112_{\pm 0.017}$
		Far	$1.148_{\pm 0.042}$	$0.134_{\pm 0.018}$	$0.108_{\pm 0.012}$	$1.301_{\pm 0.031}$	$0.141_{\pm 0.016}$	$0.115_{\pm 0.013}$	$1.389_{\pm 0.039}$	$0.146_{\pm 0.012}$	$0.112_{\pm 0.016}$
MLLS		Near	$1.152_{\pm 0.101}$	$0.131_{\pm 0.019}$	$0.099_{\pm 0.017}$	$1.233_{\pm 0.128}$	$0.143_{\pm 0.029}$	$0.111_{\pm 0.023}$	$1.272_{\pm 0.128}$	$0.146_{\pm 0.025}$	$0.116_{\pm 0.027}$
		Far	$4.095_{\pm 0.078}$	$0.167_{\pm 0.031}$	$0.101_{\pm 0.017}$	$4.270_{\pm 0.265}$	$0.182_{\pm 0.035}$	$0.112_{\pm 0.023}$	$4.436_{\pm 0.263}$	$0.189_{\pm 0.031}$	$0.117_{\pm 0.026}$
RLLS		Near	$0.432_{\pm 0.000}$	$0.432_{\pm 0.000}$	$0.432_{\pm 0.000}$	$1.083_{\pm 0.000}$	$1.082_{\pm 0.000}$	$1.082_{\pm 0.000}$	$1.397_{\pm 0.000}$	$1.396_{\pm 0.000}$	$1.396_{\pm 0.000}$
		Far	$0.433_{\pm 0.000}$	$0.432_{\pm 0.000}$	$0.432_{\pm 0.000}$	$1.083_{\pm 0.000}$	$1.082_{\pm 0.000}$	$1.082_{\pm 0.000}$	$1.397_{\pm 0.000}$	$1.396_{\pm 0.000}$	$1.396_{\pm 0.000}$
MAPLS		Near	$0.877_{\pm 0.069}$	$0.114_{\pm 0.016}$	$0.085_{\pm 0.014}$	$0.986_{\pm 0.089}$	$0.128_{\pm 0.023}$	$0.093_{\pm 0.018}$	$1.046_{\pm 0.094}$	$0.134_{\pm 0.021}$	$0.095_{\pm 0.020}$
		Far	$3.004_{\pm 0.055}$	$0.139_{\pm 0.024}$	$0.086_{\pm 0.014}$	$3.177_{\pm 0.173}$	$0.156_{\pm 0.026}$	$0.094_{\pm 0.018}$	$3.328_{\pm 0.168}$	$0.164_{\pm 0.025}$	$0.097_{\pm 0.020}$
Open Set Label Shift estimation models											
Baseline			$0.436_{\pm 0.000}$	$0.436_{\pm 0.000}$	$0.436_{\pm 0.000}$	$1.090_{\pm 0.000}$	$1.090_{\pm 0.000}$	$1.090_{\pm 0.000}$	$1.405_{\pm 0.000}$	$1.405_{\pm 0.000}$	$1.405_{\pm 0.000}$
ours	OpenMax	Near	$0.699_{\pm 0.010}$	$\mathbf{0.035}_{\pm 0.002}$	$\mathbf{0.022}_{\pm 0.002}$	$0.769_{\pm 0.015}$	$\mathbf{0.035}_{\pm 0.002}$	$\mathbf{0.020}_{\pm 0.001}$	$0.820_{\pm 0.009}$	$\mathbf{0.035}_{\pm 0.002}$	$\mathbf{0.019}_{\pm 0.000}$
		Far	$2.500_{\pm 0.153}$	$\mathbf{0.048}_{\pm 0.002}$	$\mathbf{0.022}_{\pm 0.003}$	$2.652_{\pm 0.106}$	$\mathbf{0.054}_{\pm 0.004}$	$\mathbf{0.020}_{\pm 0.001}$	$2.739_{\pm 0.114}$	$\mathbf{0.047}_{\pm 0.001}$	$\mathbf{0.019}_{\pm 0.000}$
	MLS	Near	$\mathbf{0.194}_{\pm 0.011}$	$\mathbf{0.069}_{\pm 0.005}$	$\mathbf{0.069}_{\pm 0.012}$	$\mathbf{0.209}_{\pm 0.004}$	$\mathbf{0.078}_{\pm 0.009}$	$\mathbf{0.076}_{\pm 0.009}$	$\mathbf{0.217}_{\pm 0.002}$	$\mathbf{0.079}_{\pm 0.009}$	$\mathbf{0.081}_{\pm 0.006}$
		Far	$\mathbf{0.118}_{\pm 0.024}$	$\mathbf{0.069}_{\pm 0.005}$	$\mathbf{0.069}_{\pm 0.012}$	$\mathbf{0.128}_{\pm 0.007}$	$\mathbf{0.078}_{\pm 0.009}$	$\mathbf{0.076}_{\pm 0.009}$	$\mathbf{0.126}_{\pm 0.007}$	$\mathbf{0.081}_{\pm 0.009}$	$\mathbf{0.082}_{\pm 0.006}$
	ReAct	Near	$\mathbf{0.251}_{\pm 0.050}$	$\mathbf{0.092}_{\pm 0.028}$	$0.094_{\pm 0.022}$	$\mathbf{0.253}_{\pm 0.048}$	$\mathbf{0.102}_{\pm 0.029}$	$0.097_{\pm 0.023}$	$\mathbf{0.279}_{\pm 0.052}$	$\mathbf{0.110}_{\pm 0.030}$	$0.111_{\pm 0.026}$
		Far	$\mathbf{0.103}_{\pm 0.015}$	$\mathbf{0.092}_{\pm 0.026}$	$0.094_{\pm 0.022}$	$\mathbf{0.116}_{\pm 0.017}$	$\mathbf{0.104}_{\pm 0.028}$	$0.098_{\pm 0.023}$	$\mathbf{0.128}_{\pm 0.022}$	$\mathbf{0.113}_{\pm 0.031}$	$0.112_{\pm 0.027}$
	KNN	Near	$\mathbf{0.309}_{\pm 0.008}$	$0.116_{\pm 0.007}$	$0.115_{\pm 0.005}$	$\mathbf{0.319}_{\pm 0.004}$	$0.125_{\pm 0.008}$	$0.118_{\pm 0.007}$	$\mathbf{0.325}_{\pm 0.010}$	$0.133_{\pm 0.010}$	$0.127_{\pm 0.004}$
		Far	$\mathbf{0.158}_{\pm 0.013}$	$\mathbf{0.115}_{\pm 0.007}$	$0.115_{\pm 0.005}$	$\mathbf{0.161}_{\pm 0.003}$	$\mathbf{0.125}_{\pm 0.008}$	$0.119_{\pm 0.007}$	$\mathbf{0.167}_{\pm 0.012}$	$\mathbf{0.132}_{\pm 0.010}$	$0.127_{\pm 0.003}$
	Ash	Near	$\mathbf{0.262}_{\pm 0.026}$	$\mathbf{0.110}_{\pm 0.016}$	$0.108_{\pm 0.013}$	$\mathbf{0.287}_{\pm 0.024}$	$\mathbf{0.121}_{\pm 0.015}$	$0.122_{\pm 0.006}$	$\mathbf{0.299}_{\pm 0.028}$	$\mathbf{0.132}_{\pm 0.015}$	$0.135_{\pm 0.018}$
		Far	$\mathbf{0.110}_{\pm 0.009}$	$\mathbf{0.110}_{\pm 0.016}$	$0.108_{\pm 0.013}$	$\mathbf{0.131}_{\pm 0.015}$	$\mathbf{0.121}_{\pm 0.015}$	$0.122_{\pm 0.007}$	$\mathbf{0.140}_{\pm 0.015}$	$\mathbf{0.133}_{\pm 0.014}$	$0.136_{\pm 0.018}$

Table 21. **Estimation Error $(w-\hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the ImageNet-200 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Forward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

Dataset			ImageNet-200								
ID label Shift param			LT10 Backward			LT50 Backward			LT100 Backward		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models											
BBSE		Near	$0.651_{\pm 0.079}$	$0.240_{\pm 0.077}$	$0.215_{\pm 0.049}$	$0.848_{\pm 0.074}$	$0.332_{\pm 0.086}$	$0.323_{\pm 0.085}$	$0.930_{\pm 0.063}$	$0.364_{\pm 0.112}$	$0.336_{\pm 0.090}$
		Far	$1.088_{\pm 0.049}$	$0.253_{\pm 0.055}$	$0.214_{\pm 0.046}$	$1.237_{\pm 0.054}$	$0.342_{\pm 0.075}$	$0.322_{\pm 0.085}$	$1.329_{\pm 0.045}$	$0.374_{\pm 0.096}$	$0.338_{\pm 0.090}$
MLLS		Near	$1.297_{\pm 0.163}$	$0.272_{\pm 0.086}$	$0.220_{\pm 0.043}$	$1.448_{\pm 0.143}$	$0.381_{\pm 0.099}$	$0.346_{\pm 0.092}$	$1.520_{\pm 0.135}$	$0.406_{\pm 0.111}$	$0.346_{\pm 0.081}$
		Far	$4.049_{\pm 0.286}$	$0.309_{\pm 0.057}$	$0.217_{\pm 0.040}$	$4.208_{\pm 0.222}$	$0.410_{\pm 0.093}$	$0.346_{\pm 0.089}$	$4.312_{\pm 0.173}$	$0.440_{\pm 0.093}$	$0.348_{\pm 0.081}$
RLLS		Near	$0.438_{\pm 0.000}$	$0.438_{\pm 0.000}$	$0.438_{\pm 0.000}$	$1.088_{\pm 0.000}$	$1.088_{\pm 0.000}$	$1.087_{\pm 0.000}$	$1.401_{\pm 0.000}$	$1.400_{\pm 0.000}$	$1.400_{\pm 0.000}$
		Far	$0.438_{\pm 0.000}$	$0.438_{\pm 0.000}$	$0.438_{\pm 0.000}$	$1.088_{\pm 0.000}$	$1.088_{\pm 0.000}$	$1.087_{\pm 0.000}$	$1.401_{\pm 0.000}$	$1.400_{\pm 0.000}$	$1.400_{\pm 0.000}$
MAPLS		Near	$0.992_{\pm 0.122}$	$0.229_{\pm 0.067}$	$0.187_{\pm 0.035}$	$1.170_{\pm 0.108}$	$0.331_{\pm 0.079}$	$0.295_{\pm 0.073}$	$1.253_{\pm 0.104}$	$0.357_{\pm 0.088}$	$0.298_{\pm 0.066}$
		Far	$2.930_{\pm 0.214}$	$0.249_{\pm 0.046}$	$0.184_{\pm 0.033}$	$3.063_{\pm 0.155}$	$0.340_{\pm 0.074}$	$0.294_{\pm 0.071}$	$3.170_{\pm 0.128}$	$0.369_{\pm 0.075}$	$0.298_{\pm 0.066}$
Open Set Label Shift estimation models											
Baseline			$0.435_{\pm 0.000}$	$0.435_{\pm 0.000}$	$0.435_{\pm 0.000}$	$1.086_{\pm 0.000}$	$1.086_{\pm 0.000}$	$1.086_{\pm 0.000}$	$1.400_{\pm 0.000}$	$1.400_{\pm 0.000}$	$1.400_{\pm 0.000}$
ours	OpenMax	Near	$0.739_{\pm 0.032}$	$\mathbf{0.041}_{\pm 0.000}$	$\mathbf{0.030}_{\pm 0.006}$	$0.866_{\pm 0.036}$	$\mathbf{0.050}_{\pm 0.004}$	$\mathbf{0.029}_{\pm 0.004}$	$0.940_{\pm 0.026}$	$\mathbf{0.049}_{\pm 0.001}$	$\mathbf{0.031}_{\pm 0.002}$
		Far	$2.316_{\pm 0.111}$	$\mathbf{0.060}_{\pm 0.004}$	$\mathbf{0.031}_{\pm 0.006}$	$2.466_{\pm 0.157}$	$\mathbf{0.075}_{\pm 0.002}$	$\mathbf{0.029}_{\pm 0.003}$	$2.524_{\pm 0.182}$	$\mathbf{0.075}_{\pm 0.005}$	$\mathbf{0.032}_{\pm 0.002}$
	MLS	Near	$\mathbf{0.259}_{\pm 0.010}$	$\mathbf{0.112}_{\pm 0.012}$	$\mathbf{0.114}_{\pm 0.003}$	$\mathbf{0.305}_{\pm 0.006}$	$\mathbf{0.128}_{\pm 0.012}$	$\mathbf{0.128}_{\pm 0.009}$	$\mathbf{0.328}_{\pm 0.008}$	$\mathbf{0.144}_{\pm 0.009}$	$\mathbf{0.133}_{\pm 0.013}$
		Far	$\mathbf{0.150}_{\pm 0.013}$	$\mathbf{0.108}_{\pm 0.012}$	$\mathbf{0.113}_{\pm 0.003}$	$\mathbf{0.168}_{\pm 0.003}$	$\mathbf{0.126}_{\pm 0.012}$	$\mathbf{0.129}_{\pm 0.008}$	$\mathbf{0.184}_{\pm 0.022}$	$\mathbf{0.144}_{\pm 0.009}$	$\mathbf{0.133}_{\pm 0.014}$
	ReAct	Near	$\mathbf{0.429}_{\pm 0.121}$	$\mathbf{0.199}_{\pm 0.065}$	$\mathbf{0.185}_{\pm 0.062}$	$\mathbf{0.511}_{\pm 0.167}$	$\mathbf{0.265}_{\pm 0.109}$	$\mathbf{0.255}_{\pm 0.114}$	$\mathbf{0.576}_{\pm 0.205}$	$\mathbf{0.272}_{\pm 0.118}$	$\mathbf{0.279}_{\pm 0.124}$
		Far	$\mathbf{0.210}_{\pm 0.065}$	$\mathbf{0.191}_{\pm 0.061}$	$\mathbf{0.184}_{\pm 0.062}$	$\mathbf{0.265}_{\pm 0.109}$	$\mathbf{0.261}_{\pm 0.113}$	$\mathbf{0.255}_{\pm 0.114}$	$\mathbf{0.310}_{\pm 0.141}$	$\mathbf{0.270}_{\pm 0.116}$	$\mathbf{0.279}_{\pm 0.124}$
	KNN	Near	$0.458_{\pm 0.023}$	$\mathbf{0.208}_{\pm 0.005}$	$0.205_{\pm 0.010}$	$\mathbf{0.498}_{\pm 0.006}$	$\mathbf{0.242}_{\pm 0.006}$	$\mathbf{0.247}_{\pm 0.006}$	$\mathbf{0.537}_{\pm 0.019}$	$\mathbf{0.256}_{\pm 0.014}$	$\mathbf{0.262}_{\pm 0.013}$
		Far	$\mathbf{0.273}_{\pm 0.041}$	$\mathbf{0.203}_{\pm 0.004}$	$0.204_{\pm 0.010}$	$\mathbf{0.306}_{\pm 0.032}$	$\mathbf{0.240}_{\pm 0.004}$	$\mathbf{0.247}_{\pm 0.005}$	$\mathbf{0.328}_{\pm 0.040}$	$\mathbf{0.254}_{\pm 0.015}$	$\mathbf{0.263}_{\pm 0.013}$
	Ash	Near	$\mathbf{0.401}_{\pm 0.079}$	$\mathbf{0.202}_{\pm 0.036}$	$0.196_{\pm 0.039}$	$\mathbf{0.495}_{\pm 0.086}$	$\mathbf{0.252}_{\pm 0.061}$	$\mathbf{0.261}_{\pm 0.058}$	$\mathbf{0.551}_{\pm 0.114}$	$\mathbf{0.288}_{\pm 0.070}$	$\mathbf{0.285}_{\pm 0.091}$
		Far	$\mathbf{0.197}_{\pm 0.037}$	$\mathbf{0.198}_{\pm 0.034}$	$0.195_{\pm 0.039}$	$\mathbf{0.262}_{\pm 0.058}$	$\mathbf{0.248}_{\pm 0.060}$	$\mathbf{0.260}_{\pm 0.058}$	$\mathbf{0.298}_{\pm 0.080}$	$\mathbf{0.284}_{\pm 0.069}$	$\mathbf{0.286}_{\pm 0.091}$

Table 22. **Estimation Error $(w-\hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the ImageNet-200 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Backward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

Dataset			ImageNet-200					
ID label Shift param			Dir $\alpha = 1.0$			Dir $\alpha = 10.0$		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01
Closed Set Label Shift estimation models								
BBSE		Near	$0.808_{\pm 0.125}$	$0.211_{\pm 0.044}$	$0.190_{\pm 0.045}$	$0.556_{\pm 0.048}$	$0.152_{\pm 0.038}$	$0.123_{\pm 0.015}$
		Far	$1.330_{\pm 0.085}$	$0.230_{\pm 0.042}$	$0.192_{\pm 0.045}$	$1.051_{\pm 0.065}$	$0.180_{\pm 0.039}$	$0.122_{\pm 0.015}$
MLLS		Near	$1.372_{\pm 0.224}$	$0.140_{\pm 0.026}$	$0.152_{\pm 0.061}$	$1.213_{\pm 0.168}$	$0.161_{\pm 0.047}$	$0.122_{\pm 0.028}$
		Far	$4.343_{\pm 0.354}$	$0.180_{\pm 0.041}$	$0.155_{\pm 0.062}$	$4.027_{\pm 0.316}$	$0.205_{\pm 0.049}$	$0.121_{\pm 0.028}$
RLLS		Near	$1.105_{\pm 0.091}$	$1.013_{\pm 0.052}$	$0.949_{\pm 0.094}$	$0.121_{\pm 0.011}$	$0.107_{\pm 0.007}$	$0.110_{\pm 0.009}$
		Far	$1.105_{\pm 0.091}$	$1.013_{\pm 0.052}$	$0.949_{\pm 0.094}$	$0.121_{\pm 0.011}$	$0.107_{\pm 0.007}$	$0.110_{\pm 0.009}$
MAPLS		Near	$1.129_{\pm 0.181}$	$0.148_{\pm 0.028}$	$0.126_{\pm 0.034}$	$0.886_{\pm 0.116}$	$0.133_{\pm 0.034}$	$0.101_{\pm 0.023}$
		Far	$3.228_{\pm 0.288}$	$0.171_{\pm 0.037}$	$0.128_{\pm 0.035}$	$2.900_{\pm 0.218}$	$0.161_{\pm 0.035}$	$0.100_{\pm 0.023}$
Open Set Label Shift estimation models								
Baseline			$1.100_{\pm 0.085}$	$1.013_{\pm 0.054}$	$0.948_{\pm 0.095}$	$0.122_{\pm 0.011}$	$0.108_{\pm 0.007}$	$0.110_{\pm 0.009}$
ours	OpenMax	Near	$0.764_{\pm 0.030}$	$\mathbf{0.040}_{\pm 0.010}$	$\mathbf{0.029}_{\pm 0.005}$	$0.667_{\pm 0.015}$	$\mathbf{0.046}_{\pm 0.004}$	$\mathbf{0.030}_{\pm 0.006}$
		Far	$2.456_{\pm 0.260}$	$\mathbf{0.054}_{\pm 0.009}$	$\mathbf{0.030}_{\pm 0.005}$	$2.289_{\pm 0.113}$	$\mathbf{0.060}_{\pm 0.005}$	$\mathbf{0.030}_{\pm 0.006}$
	MLS	Near	$\mathbf{0.262}_{\pm 0.008}$	$\mathbf{0.135}_{\pm 0.032}$	$\mathbf{0.132}_{\pm 0.036}$	$0.250_{\pm 0.005}$	$\mathbf{0.101}_{\pm 0.005}$	$\mathbf{0.110}_{\pm 0.007}$
		Far	$\mathbf{0.144}_{\pm 0.008}$	$\mathbf{0.132}_{\pm 0.032}$	$\mathbf{0.132}_{\pm 0.036}$	$0.151_{\pm 0.006}$	$\mathbf{0.100}_{\pm 0.004}$	$\mathbf{0.110}_{\pm 0.006}$
	ReAct	Near	$\mathbf{0.394}_{\pm 0.100}$	$0.203_{\pm 0.011}$	$0.225_{\pm 0.031}$	$0.309_{\pm 0.070}$	$0.162_{\pm 0.048}$	$0.154_{\pm 0.040}$
		Far	$\mathbf{0.234}_{\pm 0.084}$	$\mathbf{0.204}_{\pm 0.012}$	$0.225_{\pm 0.030}$	$0.155_{\pm 0.036}$	$0.157_{\pm 0.048}$	$0.155_{\pm 0.041}$
	KNN	Near	$\mathbf{0.500}_{\pm 0.051}$	$0.307_{\pm 0.129}$	$0.282_{\pm 0.103}$	$0.393_{\pm 0.016}$	$0.194_{\pm 0.021}$	$0.157_{\pm 0.004}$
		Far	$\mathbf{0.348}_{\pm 0.067}$	$0.308_{\pm 0.128}$	$0.281_{\pm 0.103}$	$0.231_{\pm 0.022}$	$0.190_{\pm 0.019}$	$0.157_{\pm 0.004}$
	Ash	Near	$\mathbf{0.415}_{\pm 0.075}$	$0.184_{\pm 0.051}$	$0.194_{\pm 0.039}$	$0.337_{\pm 0.021}$	$0.149_{\pm 0.012}$	$0.148_{\pm 0.025}$
		Far	$\mathbf{0.200}_{\pm 0.041}$	$0.183_{\pm 0.052}$	$0.194_{\pm 0.039}$	$0.146_{\pm 0.023}$	$0.150_{\pm 0.012}$	$0.147_{\pm 0.025}$

Table 23. **Estimation Error $(w - \hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the ImageNet200 dataset with Near OOD datasets and Far OOD datasets comparison under Dirichlet ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

G. More Accuracy Results

We provide the full open set label shift Top1 Accuracy on the CIFAR10 and CIFAR100 datasets. Similar to the estimation error metric, our model outperforms baseline in most of the experimental setup.

G.1. CIFAR10

ID label Shift param			LT-10			LT-50			LT 100		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
OpenMax	Near	Original	$74.650_{\pm 1.220}$	$86.800_{\pm 0.020}$	$89.690_{\pm 0.340}$	$74.340_{\pm 0.530}$	$87.650_{\pm 0.080}$	$89.940_{\pm 0.410}$	$74.750_{\pm 0.910}$	$87.160_{\pm 0.220}$	$90.190_{\pm 0.440}$
		Baseline	$75.330_{\pm 0.830}$	$86.580_{\pm 0.140}$	$89.340_{\pm 0.140}$	$75.060_{\pm 0.310}$	$87.340_{\pm 0.070}$	$89.540_{\pm 0.270}$	$75.300_{\pm 0.550}$	$86.930_{\pm 0.360}$	$89.800_{\pm 0.340}$
		ours	$75.330_{\pm 0.830}$	$86.580_{\pm 0.140}$	$89.340_{\pm 0.140}$	$75.060_{\pm 0.310}$	$87.390_{\pm 0.090}$	$89.600_{\pm 0.310}$	$75.300_{\pm 0.550}$	$86.980_{\pm 0.410}$	$89.930_{\pm 0.400}$
	Far	Original	$76.560_{\pm 0.560}$	$87.110_{\pm 0.080}$	$89.720_{\pm 0.390}$	$76.550_{\pm 0.120}$	$87.940_{\pm 0.170}$	$89.910_{\pm 0.460}$	$76.730_{\pm 0.240}$	$87.530_{\pm 0.340}$	$90.240_{\pm 0.460}$
		Baseline	$77.390_{\pm 0.030}$	$86.910_{\pm 0.070}$	$89.380_{\pm 0.190}$	$77.400_{\pm 0.380}$	$87.700_{\pm 0.090}$	$89.530_{\pm 0.320}$	$77.570_{\pm 0.320}$	$87.330_{\pm 0.470}$	$89.850_{\pm 0.350}$
		ours	$77.390_{\pm 0.030}$	$86.910_{\pm 0.070}$	$89.380_{\pm 0.190}$	$77.410_{\pm 0.360}$	$87.750_{\pm 0.130}$	$89.590_{\pm 0.360}$	$77.610_{\pm 0.320}$	$87.410_{\pm 0.480}$	$89.980_{\pm 0.410}$
MLS	Near	Baseline	$80.920_{\pm 0.260}$	$79.440_{\pm 2.360}$	$79.560_{\pm 3.030}$	$80.470_{\pm 0.380}$	$80.350_{\pm 2.910}$	$79.600_{\pm 3.360}$	$80.900_{\pm 0.880}$	$79.850_{\pm 2.680}$	$79.640_{\pm 2.770}$
		ours	$80.920_{\pm 0.260}$	$79.440_{\pm 2.360}$	$79.560_{\pm 3.030}$	$80.470_{\pm 0.380}$	$80.360_{\pm 2.920}$	$79.610_{\pm 3.370}$	$80.900_{\pm 0.880}$	$79.900_{\pm 2.680}$	$79.730_{\pm 2.800}$
	Far	Baseline	$83.900_{\pm 0.870}$	$80.040_{\pm 2.390}$	$79.620_{\pm 3.040}$	$83.580_{\pm 0.930}$	$80.810_{\pm 2.900}$	$79.710_{\pm 3.320}$	$84.020_{\pm 0.830}$	$80.260_{\pm 2.890}$	$79.760_{\pm 2.760}$
		ours	$83.900_{\pm 0.870}$	$80.040_{\pm 2.390}$	$79.620_{\pm 3.040}$	$83.590_{\pm 0.930}$	$80.820_{\pm 2.910}$	$79.720_{\pm 3.330}$	$84.040_{\pm 0.840}$	$80.310_{\pm 2.890}$	$79.850_{\pm 2.790}$
ReAct	Near	Baseline	$80.120_{\pm 0.760}$	$78.450_{\pm 3.120}$	$78.570_{\pm 3.590}$	$79.330_{\pm 0.520}$	$77.720_{\pm 3.110}$	$77.710_{\pm 4.050}$	$79.170_{\pm 0.640}$	$77.040_{\pm 2.380}$	$76.980_{\pm 3.380}$
		ours	$80.120_{\pm 0.760}$	$78.460_{\pm 3.120}$	$78.570_{\pm 3.590}$	$79.320_{\pm 0.520}$	$77.720_{\pm 3.110}$	$77.710_{\pm 4.050}$	$79.170_{\pm 0.640}$	$77.040_{\pm 2.380}$	$76.980_{\pm 3.380}$
	Far	Baseline	$82.620_{\pm 1.460}$	$79.040_{\pm 3.020}$	$78.600_{\pm 3.590}$	$82.120_{\pm 1.140}$	$78.020_{\pm 2.910}$	$77.780_{\pm 4.050}$	$81.570_{\pm 1.330}$	$77.610_{\pm 2.360}$	$76.980_{\pm 3.390}$
		ours	$82.620_{\pm 1.460}$	$79.050_{\pm 3.020}$	$78.600_{\pm 3.590}$	$82.120_{\pm 1.130}$	$78.020_{\pm 2.910}$	$77.780_{\pm 4.050}$	$81.570_{\pm 1.330}$	$77.610_{\pm 2.360}$	$76.980_{\pm 3.390}$
KNN	Near	Baseline	$80.760_{\pm 0.800}$	$84.740_{\pm 0.870}$	$85.730_{\pm 1.350}$	$80.970_{\pm 1.130}$	$84.640_{\pm 0.910}$	$85.710_{\pm 1.660}$	$81.120_{\pm 0.820}$	$85.030_{\pm 0.790}$	$85.820_{\pm 1.280}$
		ours	$80.760_{\pm 0.800}$	$84.740_{\pm 0.870}$	$85.730_{\pm 1.350}$	$80.970_{\pm 1.130}$	$84.670_{\pm 0.930}$	$85.750_{\pm 1.700}$	$81.120_{\pm 0.820}$	$85.100_{\pm 0.780}$	$85.950_{\pm 1.260}$
	Far	Baseline	$84.030_{\pm 0.840}$	$85.230_{\pm 0.780}$	$85.760_{\pm 1.330}$	$84.160_{\pm 0.800}$	$85.460_{\pm 0.910}$	$85.680_{\pm 1.680}$	$84.460_{\pm 0.660}$	$85.650_{\pm 0.760}$	$85.850_{\pm 1.300}$
		ours	$84.030_{\pm 0.840}$	$85.230_{\pm 0.780}$	$85.760_{\pm 1.330}$	$84.180_{\pm 0.790}$	$85.490_{\pm 0.930}$	$85.720_{\pm 1.710}$	$84.490_{\pm 0.670}$	$85.740_{\pm 0.740}$	$85.980_{\pm 1.280}$
Ash	Near	Baseline	$68.480_{\pm 2.030}$	$68.570_{\pm 4.160}$	$69.700_{\pm 4.480}$	$68.190_{\pm 2.750}$	$68.650_{\pm 4.000}$	$68.990_{\pm 4.880}$	$67.950_{\pm 2.720}$	$68.770_{\pm 4.260}$	$69.290_{\pm 5.380}$
		ours	$68.490_{\pm 2.040}$	$68.580_{\pm 4.160}$	$69.700_{\pm 4.460}$	$68.210_{\pm 2.740}$	$68.660_{\pm 3.990}$	$69.010_{\pm 4.880}$	$67.960_{\pm 2.710}$	$68.840_{\pm 4.200}$	$69.400_{\pm 5.300}$
	Far	Baseline	$70.900_{\pm 2.810}$	$68.980_{\pm 4.010}$	$69.730_{\pm 4.490}$	$70.770_{\pm 3.200}$	$69.100_{\pm 4.400}$	$69.120_{\pm 4.930}$	$70.330_{\pm 3.280}$	$69.410_{\pm 4.280}$	$69.280_{\pm 5.340}$
		ours	$70.910_{\pm 2.820}$	$69.000_{\pm 4.010}$	$69.730_{\pm 4.480}$	$70.780_{\pm 3.200}$	$69.110_{\pm 4.380}$	$69.150_{\pm 4.930}$	$70.360_{\pm 3.270}$	$69.500_{\pm 4.220}$	$69.400_{\pm 5.260}$

Table 24. **Top1 Accuracy (↑) of our OSLS correction model on the CIFAR10 dataset with Near OOD datasets and Far OOD datasets comparison under under Ordered-LT (Backward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

ID label Shift param			LT-10			LT-50			LT 100		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
OpenMax	Near	Original	$75.010_{\pm 0.970}$	$88.340_{\pm 0.440}$	$90.400_{\pm 0.410}$	$75.610_{\pm 1.210}$	$88.670_{\pm 0.320}$	$91.100_{\pm 0.330}$	$75.430_{\pm 1.280}$	$88.860_{\pm 0.160}$	$91.260_{\pm 0.060}$
		Baseline	$75.720_{\pm 0.660}$	$88.210_{\pm 0.480}$	$90.100_{\pm 0.310}$	$76.300_{\pm 0.870}$	$88.610_{\pm 0.380}$	$90.860_{\pm 0.370}$	$76.170_{\pm 0.940}$	$88.830_{\pm 0.230}$	$91.070_{\pm 0.030}$
		ours	$75.720_{\pm 0.660}$	$88.200_{\pm 0.480}$	$90.110_{\pm 0.300}$	$76.300_{\pm 0.880}$	$88.650_{\pm 0.370}$	$90.950_{\pm 0.350}$	$76.180_{\pm 0.950}$	$88.950_{\pm 0.230}$	$91.200_{\pm 0.040}$
	Far	Original	$76.930_{\pm 0.520}$	$88.770_{\pm 0.350}$	$90.450_{\pm 0.340}$	$77.750_{\pm 0.650}$	$88.870_{\pm 0.370}$	$91.070_{\pm 0.240}$	$77.550_{\pm 0.610}$	$89.260_{\pm 0.160}$	$91.360_{\pm 0.060}$
		Baseline	$77.820_{\pm 0.200}$	$88.670_{\pm 0.400}$	$90.150_{\pm 0.260}$	$78.590_{\pm 0.300}$	$88.860_{\pm 0.430}$	$90.860_{\pm 0.280}$	$78.480_{\pm 0.300}$	$89.230_{\pm 0.280}$	$91.190_{\pm 0.020}$
		ours	$77.820_{\pm 0.200}$	$88.660_{\pm 0.400}$	$90.160_{\pm 0.250}$	$78.600_{\pm 0.310}$	$88.900_{\pm 0.410}$	$90.950_{\pm 0.250}$	$78.510_{\pm 0.310}$	$89.350_{\pm 0.260}$	$91.310_{\pm 0.040}$
MLS	Near	Baseline	$82.380_{\pm 0.530}$	$82.120_{\pm 1.540}$	$81.810_{\pm 2.250}$	$83.000_{\pm 0.250}$	$83.180_{\pm 1.860}$	$83.180_{\pm 2.320}$	$82.770_{\pm 0.070}$	$83.670_{\pm 2.170}$	$83.640_{\pm 2.650}$
		ours	$82.380_{\pm 0.530}$	$82.120_{\pm 1.540}$	$81.810_{\pm 2.250}$	$83.000_{\pm 0.250}$	$83.180_{\pm 1.860}$	$83.180_{\pm 2.320}$	$82.770_{\pm 0.070}$	$83.680_{\pm 2.170}$	$83.640_{\pm 2.650}$
	Far	Baseline	$85.410_{\pm 0.550}$	$82.660_{\pm 1.690}$	$81.840_{\pm 2.310}$	$85.870_{\pm 0.420}$	$83.780_{\pm 1.960}$	$83.250_{\pm 2.290}$	$85.890_{\pm 0.580}$	$84.190_{\pm 2.310}$	$83.680_{\pm 2.620}$
		ours	$85.410_{\pm 0.550}$	$82.660_{\pm 1.690}$	$81.840_{\pm 2.310}$	$85.870_{\pm 0.420}$	$83.780_{\pm 1.960}$	$83.250_{\pm 2.290}$	$85.890_{\pm 0.580}$	$84.200_{\pm 2.310}$	$83.680_{\pm 2.620}$
ReAct	Near	Baseline	$81.730_{\pm 0.620}$	$81.550_{\pm 0.790}$	$81.970_{\pm 1.330}$	$82.310_{\pm 0.970}$	$82.210_{\pm 1.140}$	$81.670_{\pm 1.940}$	$81.960_{\pm 1.060}$	$82.090_{\pm 1.470}$	$82.000_{\pm 1.620}$
		ours	$81.730_{\pm 0.620}$	$81.550_{\pm 0.790}$	$81.970_{\pm 1.330}$	$82.310_{\pm 0.970}$	$82.210_{\pm 1.140}$	$81.650_{\pm 1.930}$	$81.960_{\pm 1.060}$	$82.090_{\pm 1.470}$	$82.000_{\pm 1.620}$
	Far	Baseline	$84.360_{\pm 1.280}$	$82.060_{\pm 0.690}$	$82.040_{\pm 1.280}$	$84.830_{\pm 1.250}$	$82.720_{\pm 1.120}$	$81.720_{\pm 1.860}$	$84.420_{\pm 1.580}$	$82.470_{\pm 1.320}$	$82.000_{\pm 1.580}$
		ours	$84.360_{\pm 1.280}$	$82.060_{\pm 0.690}$	$82.040_{\pm 1.280}$	$84.830_{\pm 1.250}$	$82.720_{\pm 1.120}$	$81.700_{\pm 1.840}$	$84.420_{\pm 1.580}$	$82.470_{\pm 1.320}$	$82.020_{\pm 1.580}$
KNN	Near	Baseline	$81.840_{\pm 0.680}$	$87.030_{\pm 0.430}$	$88.050_{\pm 1.100}$	$82.490_{\pm 1.280}$	$87.920_{\pm 1.060}$	$88.880_{\pm 0.980}$	$82.290_{\pm 0.900}$	$88.590_{\pm 0.600}$	$89.400_{\pm 0.730}$
		ours	$81.840_{\pm 0.680}$	$87.040_{\pm 0.420}$	$88.060_{\pm 1.090}$	$82.520_{\pm 1.290}$	$88.000_{\pm 1.030}$	$88.960_{\pm 0.960}$	$82.350_{\pm 0.910}$	$88.730_{\pm 0.590}$	$89.590_{\pm 0.770}$
	Far	Baseline	$84.960_{\pm 0.720}$	$87.570_{\pm 0.540}$	$88.080_{\pm 1.070}$	$85.670_{\pm 0.690}$	$88.550_{\pm 0.820}$	$88.930_{\pm 0.960}$	$85.480_{\pm 0.910}$	$89.150_{\pm 0.640}$	$89.460_{\pm 0.720}$
		ours	$84.960_{\pm 0.730}$	$87.580_{\pm 0.530}$	$88.090_{\pm 1.060}$	$85.710_{\pm 0.700}$	$88.640_{\pm 0.800}$	$89.020_{\pm 0.930}$	$85.550_{\pm 0.910}$	$89.290_{\pm 0.620}$	$89.650_{\pm 0.760}$
Ash	Near	Baseline	$66.860_{\pm 4.450}$	$64.980_{\pm 7.700}$	$65.040_{\pm 8.630}$	$64.510_{\pm 6.860}$	$62.000_{\pm 13.130}$	$61.080_{\pm 14.280}$	$64.110_{\pm 8.000}$	$60.340_{\pm 14.610}$	$59.600_{\pm 16.470}$
		ours	$66.860_{\pm 4.450}$	$64.990_{\pm 7.710}$	$65.040_{\pm 8.630}$	$64.510_{\pm 6.860}$	$62.000_{\pm 13.130}$	$61.090_{\pm 14.290}$	$64.120_{\pm 8.000}$	$60.340_{\pm 14.610}$	$59.600_{\pm 16.470}$
	Far	Baseline	$68.820_{\pm 4.610}$	$65.290_{\pm 7.740}$	$65.050_{\pm 8.660}$	$66.940_{\pm 6.870}$	$62.290_{\pm 12.960}$	$61.100_{\pm 14.240}$	$66.040_{\pm 8.160}$	$60.560_{\pm 14.670}$	$59.700_{\pm 16.460}$
		ours	$68.820_{\pm 4.620}$	$65.300_{\pm 7.750}$	$65.050_{\pm 8.660}$	$66.950_{\pm 6.870}$	$62.290_{\pm 12.960}$	$61.100_{\pm 14.240}$	$66.040_{\pm 8.150}$	$60.560_{\pm 14.670}$	$59.700_{\pm 16.460}$

Table 25. **Top1 Accuracy (↑) of our OSLS correction model on the CIFAR10 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Backward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

ID label Shift param			Dir $\alpha = 1.0$			Dir $\alpha = 10.0$		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01
OpenMax	Near	Original	$74.54_{\pm0.65}$	$87.11_{\pm0.57}$	$89.33_{\pm1.08}$	$74.80_{\pm1.17}$	$86.84_{\pm0.30}$	$89.85_{\pm0.61}$
		Baseline	$76.73_{\pm1.53}$	$86.45_{\pm0.54}$	$88.00_{\pm1.22}$	$76.37_{\pm1.00}$	$86.29_{\pm0.31}$	$88.65_{\pm0.99}$
		ours	$76.73_{\pm1.53}$	$86.46_{\pm0.55}$	$88.00_{\pm1.22}$	$76.37_{\pm1.00}$	$86.29_{\pm0.31}$	$88.65_{\pm0.99}$
	Far	Original	$76.35_{\pm0.39}$	$87.63_{\pm0.58}$	$89.40_{\pm1.02}$	$76.79_{\pm0.85}$	$87.11_{\pm0.20}$	$89.88_{\pm0.61}$
		Baseline	$78.91_{\pm2.30}$	$86.95_{\pm0.39}$	$88.07_{\pm1.23}$	$78.86_{\pm1.75}$	$86.63_{\pm0.35}$	$88.71_{\pm1.06}$
		ours	$78.91_{\pm2.30}$	$86.96_{\pm0.41}$	$88.07_{\pm1.22}$	$78.86_{\pm1.75}$	$86.63_{\pm0.35}$	$88.71_{\pm1.06}$
MLS	Near	Baseline	$81.35_{\pm0.25}$	$79.87_{\pm1.73}$	$81.34_{\pm3.17}$	$81.25_{\pm0.26}$	$81.00_{\pm2.30}$	$79.56_{\pm2.18}$
		ours	$81.35_{\pm0.25}$	$79.87_{\pm1.74}$	$81.34_{\pm3.17}$	$81.25_{\pm0.26}$	$81.00_{\pm2.30}$	$79.56_{\pm2.18}$
	Far	Baseline	$84.41_{\pm1.13}$	$80.47_{\pm1.80}$	$81.43_{\pm3.12}$	$84.35_{\pm0.74}$	$81.43_{\pm2.39}$	$79.61_{\pm2.21}$
		ours	$84.41_{\pm1.12}$	$80.48_{\pm1.82}$	$81.43_{\pm3.12}$	$84.35_{\pm0.74}$	$81.43_{\pm2.39}$	$79.61_{\pm2.21}$
ReAct	Near	Baseline	$80.35_{\pm0.83}$	$83.34_{\pm0.92}$	$80.99_{\pm3.18}$	$81.00_{\pm0.20}$	$80.88_{\pm2.25}$	$81.91_{\pm2.36}$
		ours	$80.35_{\pm0.83}$	$83.38_{\pm0.94}$	$81.02_{\pm3.20}$	$81.00_{\pm0.20}$	$80.88_{\pm2.25}$	$81.91_{\pm2.36}$
	Far	Baseline	$83.06_{\pm2.01}$	$83.77_{\pm0.84}$	$81.04_{\pm3.17}$	$83.62_{\pm1.09}$	$81.39_{\pm2.39}$	$81.96_{\pm2.34}$
		ours	$83.07_{\pm2.00}$	$83.82_{\pm0.87}$	$81.07_{\pm3.19}$	$83.62_{\pm1.09}$	$81.39_{\pm2.39}$	$81.96_{\pm2.34}$
KNN	Near	Baseline	$83.11_{\pm1.31}$	$82.10_{\pm3.85}$	$85.74_{\pm1.46}$	$82.87_{\pm0.27}$	$84.39_{\pm1.22}$	$83.74_{\pm1.07}$
		ours	$83.11_{\pm1.31}$	$82.10_{\pm3.85}$	$85.74_{\pm1.46}$	$82.87_{\pm0.27}$	$84.39_{\pm1.22}$	$83.74_{\pm1.07}$
	Far	Baseline	$86.05_{\pm0.91}$	$82.51_{\pm3.96}$	$85.77_{\pm1.42}$	$85.64_{\pm0.36}$	$84.92_{\pm1.20}$	$83.79_{\pm1.08}$
		ours	$86.05_{\pm0.91}$	$82.51_{\pm3.96}$	$85.77_{\pm1.42}$	$85.64_{\pm0.36}$	$84.92_{\pm1.20}$	$83.79_{\pm1.08}$
Ash	Near	Baseline	$67.19_{\pm5.01}$	$70.88_{\pm2.22}$	$65.81_{\pm8.96}$	$67.29_{\pm2.19}$	$69.72_{\pm1.81}$	$68.70_{\pm2.27}$
		ours	$67.19_{\pm5.02}$	$70.89_{\pm2.22}$	$65.83_{\pm8.99}$	$67.30_{\pm2.19}$	$69.72_{\pm1.81}$	$68.71_{\pm2.26}$
	Far	Baseline	$69.50_{\pm4.87}$	$71.33_{\pm2.11}$	$65.86_{\pm8.98}$	$69.47_{\pm2.84}$	$70.31_{\pm1.64}$	$68.69_{\pm2.17}$
		ours	$69.50_{\pm4.88}$	$71.34_{\pm2.11}$	$65.88_{\pm9.02}$	$69.48_{\pm2.83}$	$70.31_{\pm1.64}$	$68.70_{\pm2.16}$

Table 26. **Estimation Error $(w - \hat{w})^2/K(\downarrow)$ of our OSLS estimation model on the CIFAR10 dataset with Near OOD datasets and Far OOD datasets comparison under Dirichlet ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

G.2. CIFAR100

ID label Shift param			LT-10			LT-50			LT 100		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
OpenMax	Near	Original	$63.540_{\pm 0.260}$	$65.200_{\pm 0.350}$	$65.350_{\pm 0.370}$	$63.700_{\pm 0.360}$	$65.120_{\pm 0.480}$	$64.960_{\pm 0.630}$	$63.420_{\pm 0.330}$	$64.860_{\pm 0.360}$	$65.200_{\pm 0.700}$
		Baseline	$62.710_{\pm 0.170}$	$64.920_{\pm 0.320}$	$65.130_{\pm 0.300}$	$62.950_{\pm 0.350}$	$64.720_{\pm 0.550}$	$64.610_{\pm 0.640}$	$62.670_{\pm 0.270}$	$64.610_{\pm 0.280}$	$65.100_{\pm 0.520}$
		ours	$62.950_{\pm 0.200}$	$65.250_{\pm 0.340}$	$65.700_{\pm 0.400}$	$63.430_{\pm 0.410}$	$65.780_{\pm 0.510}$	$65.900_{\pm 0.400}$	$63.230_{\pm 0.370}$	$66.030_{\pm 0.680}$	$66.760_{\pm 0.880}$
	Far	Original	$66.770_{\pm 0.340}$	$65.930_{\pm 0.360}$	$65.430_{\pm 0.400}$	$67.110_{\pm 0.380}$	$65.950_{\pm 0.530}$	$65.010_{\pm 0.690}$	$66.760_{\pm 0.060}$	$65.320_{\pm 0.380}$	$65.300_{\pm 0.720}$
		Baseline	$66.580_{\pm 0.380}$	$65.730_{\pm 0.340}$	$65.230_{\pm 0.360}$	$66.990_{\pm 0.590}$	$65.690_{\pm 0.650}$	$64.670_{\pm 0.710}$	$66.580_{\pm 0.250}$	$65.190_{\pm 0.290}$	$65.200_{\pm 0.540}$
		ours	$66.780_{\pm 0.370}$	$66.030_{\pm 0.300}$	$65.800_{\pm 0.420}$	$67.370_{\pm 0.370}$	$66.730_{\pm 0.420}$	$65.940_{\pm 0.440}$	$67.240_{\pm 0.090}$	$66.670_{\pm 0.770}$	$66.880_{\pm 0.910}$
MLS	Near	Baseline	$71.220_{\pm 0.230}$	$57.550_{\pm 0.250}$	$54.360_{\pm 0.540}$	$70.920_{\pm 0.380}$	$56.740_{\pm 0.580}$	$54.610_{\pm 0.170}$	$70.560_{\pm 0.250}$	$57.120_{\pm 0.350}$	$54.900_{\pm 0.270}$
		ours	$71.250_{\pm 0.230}$	$57.700_{\pm 0.260}$	$54.470_{\pm 0.600}$	$71.090_{\pm 0.350}$	$57.150_{\pm 0.780}$	$55.090_{\pm 0.330}$	$70.810_{\pm 0.280}$	$57.560_{\pm 0.370}$	$55.350_{\pm 0.160}$
	Far	Baseline	$70.770_{\pm 0.700}$	$57.490_{\pm 0.120}$	$54.370_{\pm 0.530}$	$70.340_{\pm 0.550}$	$56.660_{\pm 0.560}$	$54.550_{\pm 0.170}$	$70.110_{\pm 0.570}$	$56.930_{\pm 0.300}$	$54.910_{\pm 0.240}$
		ours	$70.810_{\pm 0.700}$	$57.630_{\pm 0.180}$	$54.490_{\pm 0.600}$	$70.540_{\pm 0.570}$	$57.060_{\pm 0.760}$	$55.040_{\pm 0.320}$	$70.400_{\pm 0.580}$	$57.380_{\pm 0.320}$	$55.370_{\pm 0.130}$
ReAct	Near	Baseline	$70.780_{\pm 0.130}$	$57.310_{\pm 0.260}$	$54.990_{\pm 0.570}$	$70.830_{\pm 0.430}$	$56.460_{\pm 0.550}$	$54.760_{\pm 0.420}$	$70.640_{\pm 0.570}$	$57.220_{\pm 0.570}$	$54.890_{\pm 0.480}$
		ours	$70.810_{\pm 0.150}$	$57.520_{\pm 0.270}$	$55.070_{\pm 0.500}$	$70.990_{\pm 0.450}$	$56.870_{\pm 0.680}$	$55.070_{\pm 0.270}$	$70.880_{\pm 0.510}$	$57.650_{\pm 0.690}$	$55.350_{\pm 0.600}$
	Far	Baseline	$71.150_{\pm 0.140}$	$57.170_{\pm 0.350}$	$55.020_{\pm 0.570}$	$71.110_{\pm 0.780}$	$56.700_{\pm 0.400}$	$54.750_{\pm 0.490}$	$70.770_{\pm 0.940}$	$57.290_{\pm 0.490}$	$54.830_{\pm 0.490}$
		ours	$71.160_{\pm 0.090}$	$57.410_{\pm 0.340}$	$55.100_{\pm 0.490}$	$71.320_{\pm 0.740}$	$57.090_{\pm 0.480}$	$55.070_{\pm 0.310}$	$71.070_{\pm 0.870}$	$57.730_{\pm 0.630}$	$55.300_{\pm 0.610}$
KNN	Near	Baseline	$71.110_{\pm 0.540}$	$59.240_{\pm 0.530}$	$56.650_{\pm 0.600}$	$70.500_{\pm 0.660}$	$59.410_{\pm 0.190}$	$57.130_{\pm 0.370}$	$70.720_{\pm 0.220}$	$59.070_{\pm 0.210}$	$57.400_{\pm 0.320}$
		ours	$71.170_{\pm 0.540}$	$59.420_{\pm 0.640}$	$56.840_{\pm 0.580}$	$70.720_{\pm 0.570}$	$59.860_{\pm 0.260}$	$57.570_{\pm 0.280}$	$70.920_{\pm 0.310}$	$59.630_{\pm 0.430}$	$58.120_{\pm 0.480}$
	Far	Baseline	$72.220_{\pm 0.300}$	$59.470_{\pm 0.670}$	$56.710_{\pm 0.580}$	$71.820_{\pm 0.540}$	$59.890_{\pm 0.340}$	$57.140_{\pm 0.360}$	$71.760_{\pm 0.320}$	$59.110_{\pm 0.280}$	$57.410_{\pm 0.360}$
		ours	$72.280_{\pm 0.300}$	$59.640_{\pm 0.800}$	$56.890_{\pm 0.560}$	$72.040_{\pm 0.480}$	$60.340_{\pm 0.370}$	$57.590_{\pm 0.270}$	$71.970_{\pm 0.360}$	$59.690_{\pm 0.510}$	$58.150_{\pm 0.520}$
Ash	Near	Baseline	$68.320_{\pm 0.260}$	$57.000_{\pm 0.950}$	$55.030_{\pm 0.280}$	$67.770_{\pm 0.520}$	$56.040_{\pm 0.390}$	$53.860_{\pm 0.870}$	$67.890_{\pm 0.050}$	$56.300_{\pm 0.670}$	$54.030_{\pm 0.920}$
		ours	$68.370_{\pm 0.280}$	$57.090_{\pm 0.930}$	$55.060_{\pm 0.390}$	$67.950_{\pm 0.400}$	$56.400_{\pm 0.520}$	$54.070_{\pm 0.820}$	$68.110_{\pm 0.090}$	$56.840_{\pm 0.650}$	$54.570_{\pm 1.140}$
	Far	Baseline	$70.010_{\pm 0.380}$	$57.240_{\pm 0.800}$	$55.090_{\pm 0.290}$	$69.920_{\pm 1.050}$	$56.530_{\pm 0.260}$	$53.950_{\pm 0.920}$	$69.770_{\pm 0.590}$	$56.670_{\pm 0.630}$	$54.120_{\pm 0.910}$
		ours	$70.050_{\pm 0.350}$	$57.310_{\pm 0.770}$	$55.120_{\pm 0.400}$	$70.100_{\pm 0.930}$	$56.930_{\pm 0.380}$	$54.160_{\pm 0.860}$	$70.050_{\pm 0.470}$	$57.210_{\pm 0.510}$	$54.650_{\pm 1.130}$

Table 27. **Top1 Accuracy (↑) of our OSLS estimation and correction model on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Forward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

ID label Shift param			LT-10			LT-50			LT 100		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01	1.0	0.1	0.01
OpenMax	Near	Original	$63.770_{\pm 0.360}$	$65.680_{\pm 0.340}$	$65.990_{\pm 0.380}$	$63.760_{\pm 0.440}$	$66.280_{\pm 0.990}$	$66.390_{\pm 0.320}$	$63.960_{\pm 0.580}$	$65.920_{\pm 0.590}$	$66.550_{\pm 1.100}$
		Baseline	$63.000_{\pm 0.300}$	$65.540_{\pm 0.280}$	$65.480_{\pm 0.270}$	$63.030_{\pm 0.190}$	$65.920_{\pm 0.830}$	$65.890_{\pm 0.220}$	$63.290_{\pm 0.550}$	$65.490_{\pm 0.350}$	$66.240_{\pm 0.740}$
		ours	$63.210_{\pm 0.280}$	$65.960_{\pm 0.390}$	$65.990_{\pm 0.220}$	$63.680_{\pm 0.130}$	$67.050_{\pm 0.930}$	$67.200_{\pm 0.110}$	$63.930_{\pm 0.460}$	$67.080_{\pm 0.420}$	$67.910_{\pm 0.990}$
	Far	Original	$67.160_{\pm 0.970}$	$66.180_{\pm 0.350}$	$66.020_{\pm 0.330}$	$66.800_{\pm 0.780}$	$66.960_{\pm 0.700}$	$66.510_{\pm 0.340}$	$67.390_{\pm 0.950}$	$66.490_{\pm 0.720}$	$66.630_{\pm 1.070}$
		Baseline	$66.940_{\pm 0.920}$	$66.100_{\pm 0.130}$	$65.520_{\pm 0.200}$	$66.750_{\pm 0.750}$	$66.710_{\pm 0.490}$	$66.010_{\pm 0.250}$	$67.250_{\pm 0.930}$	$66.220_{\pm 0.440}$	$66.300_{\pm 0.740}$
		ours	$67.150_{\pm 0.950}$	$66.540_{\pm 0.260}$	$66.040_{\pm 0.160}$	$67.460_{\pm 0.850}$	$67.920_{\pm 0.570}$	$67.340_{\pm 0.160}$	$68.010_{\pm 0.970}$	$67.850_{\pm 0.620}$	$68.000_{\pm 0.970}$
MLS	Near	Baseline	$71.690_{\pm 0.180}$	$58.330_{\pm 0.440}$	$55.590_{\pm 0.480}$	$72.020_{\pm 0.530}$	$58.400_{\pm 0.870}$	$56.560_{\pm 0.660}$	$71.920_{\pm 0.190}$	$58.760_{\pm 0.740}$	$55.850_{\pm 0.160}$
		ours	$71.710_{\pm 0.200}$	$58.440_{\pm 0.470}$	$55.720_{\pm 0.520}$	$72.150_{\pm 0.580}$	$58.690_{\pm 0.800}$	$56.740_{\pm 0.530}$	$72.070_{\pm 0.200}$	$59.150_{\pm 0.750}$	$56.300_{\pm 0.240}$
	Far	Baseline	$71.340_{\pm 0.400}$	$58.210_{\pm 0.380}$	$55.530_{\pm 0.460}$	$71.640_{\pm 0.480}$	$58.250_{\pm 0.840}$	$56.560_{\pm 0.680}$	$71.370_{\pm 0.380}$	$58.610_{\pm 0.570}$	$55.860_{\pm 0.150}$
		ours	$71.360_{\pm 0.390}$	$58.330_{\pm 0.420}$	$55.660_{\pm 0.500}$	$71.790_{\pm 0.530}$	$58.560_{\pm 0.790}$	$56.730_{\pm 0.550}$	$71.550_{\pm 0.380}$	$59.010_{\pm 0.600}$	$56.310_{\pm 0.250}$
ReAct	Near	Baseline	$71.640_{\pm 0.480}$	$58.280_{\pm 0.960}$	$56.020_{\pm 0.130}$	$71.370_{\pm 0.190}$	$58.640_{\pm 0.830}$	$55.280_{\pm 0.240}$	$71.590_{\pm 0.250}$	$58.210_{\pm 0.900}$	$55.600_{\pm 0.590}$
		ours	$71.680_{\pm 0.520}$	$58.480_{\pm 1.060}$	$56.170_{\pm 0.130}$	$71.490_{\pm 0.130}$	$58.930_{\pm 0.860}$	$55.820_{\pm 0.270}$	$71.840_{\pm 0.360}$	$58.700_{\pm 0.970}$	$56.130_{\pm 0.730}$
	Far	Baseline	$71.790_{\pm 0.480}$	$58.380_{\pm 0.730}$	$56.000_{\pm 0.120}$	$71.380_{\pm 0.570}$	$58.510_{\pm 0.760}$	$55.300_{\pm 0.240}$	$71.690_{\pm 0.570}$	$58.130_{\pm 0.840}$	$55.630_{\pm 0.600}$
		ours	$71.860_{\pm 0.510}$	$58.600_{\pm 0.860}$	$56.140_{\pm 0.120}$	$71.560_{\pm 0.540}$	$58.800_{\pm 0.790}$	$55.850_{\pm 0.270}$	$72.000_{\pm 0.670}$	$58.640_{\pm 0.910}$	$56.160_{\pm 0.730}$
KNN	Near	Baseline	$71.470_{\pm 0.360}$	$60.480_{\pm 0.450}$	$58.830_{\pm 1.140}$	$71.570_{\pm 0.150}$	$61.520_{\pm 1.230}$	$59.080_{\pm 0.910}$	$71.430_{\pm 0.210}$	$60.790_{\pm 0.750}$	$59.270_{\pm 0.790}$
		ours	$71.520_{\pm 0.360}$	$60.570_{\pm 0.550}$	$58.950_{\pm 1.330}$	$71.750_{\pm 0.240}$	$61.850_{\pm 1.170}$	$59.470_{\pm 0.840}$	$71.730_{\pm 0.210}$	$61.400_{\pm 0.620}$	$59.890_{\pm 0.810}$
	Far	Baseline	$72.570_{\pm 0.260}$	$60.810_{\pm 0.460}$	$58.810_{\pm 1.120}$	$72.510_{\pm 0.340}$	$61.670_{\pm 1.240}$	$59.110_{\pm 1.020}$	$72.500_{\pm 0.460}$	$61.090_{\pm 0.840}$	$59.280_{\pm 0.800}$
		ours	$72.630_{\pm 0.280}$	$60.900_{\pm 0.540}$	$58.940_{\pm 1.310}$	$72.720_{\pm 0.300}$	$62.030_{\pm 1.180}$	$59.510_{\pm 0.970}$	$72.830_{\pm 0.440}$	$61.710_{\pm 0.710}$	$59.900_{\pm 0.820}$
Ash	Near	Baseline	$68.540_{\pm 0.410}$	$57.790_{\pm 1.430}$	$54.480_{\pm 1.230}$	$68.020_{\pm 1.100}$	$57.470_{\pm 2.070}$	$55.260_{\pm 1.920}$	$68.330_{\pm 0.760}$	$57.400_{\pm 1.530}$	$55.710_{\pm 1.690}$
		ours	$68.610_{\pm 0.410}$	$57.930_{\pm 1.520}$	$54.660_{\pm 1.300}$	$68.200_{\pm 1.160}$	$57.820_{\pm 2.110}$	$55.640_{\pm 1.940}$	$68.610_{\pm 0.860}$	$57.960_{\pm 1.600}$	$56.390_{\pm 1.830}$
	Far	Baseline	$70.020_{\pm 0.540}$	$58.120_{\pm 1.410}$	$54.550_{\pm 1.270}$	$70.030_{\pm 0.130}$	$57.820_{\pm 1.750}$	$55.250_{\pm 1.880}$	$70.390_{\pm 0.450}$	$57.690_{\pm 1.560}$	$55.740_{\pm 1.700}$
		ours	$70.110_{\pm 0.540}$	$58.250_{\pm 1.480}$	$54.730_{\pm 1.340}$	$70.260_{\pm 0.180}$	$58.170_{\pm 1.820}$	$55.650_{\pm 1.900}$	$70.710_{\pm 0.460}$	$58.260_{\pm 1.600}$	$56.410_{\pm 1.860}$

Table 28. **Top1 Accuracy (↑) of our OSLS estimation and correction model on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Ordered-LT (Backward) ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.

ID label Shift param			Dir $\alpha = 1.0$			Dir $\alpha = 10.0$		
OOD label shift param r			1.0	0.1	0.01	1.0	0.1	0.01
OpenMax	Near	Original	$63.77_{\pm 0.34}$	$66.25_{\pm 0.75}$	$65.58_{\pm 0.70}$	$64.18_{\pm 0.14}$	$65.16_{\pm 0.32}$	$66.96_{\pm 0.22}$
		Baseline	$66.59_{\pm 0.46}$	$64.31_{\pm 0.19}$	$62.61_{\pm 1.53}$	$66.69_{\pm 0.95}$	$63.13_{\pm 0.88}$	$64.22_{\pm 0.73}$
		ours	$66.96_{\pm 0.36}$	$65.19_{\pm 0.13}$	$63.30_{\pm 1.54}$	$66.72_{\pm 0.85}$	$63.22_{\pm 0.97}$	$64.19_{\pm 0.82}$
	Far	Original	$67.16_{\pm 0.69}$	$66.89_{\pm 0.72}$	$65.68_{\pm 0.66}$	$67.29_{\pm 0.44}$	$65.73_{\pm 0.71}$	$67.06_{\pm 0.22}$
		Baseline	$69.64_{\pm 0.94}$	$64.90_{\pm 0.16}$	$62.70_{\pm 1.51}$	$69.71_{\pm 1.06}$	$63.72_{\pm 1.17}$	$64.27_{\pm 0.72}$
		ours	$70.02_{\pm 0.77}$	$65.77_{\pm 0.16}$	$63.37_{\pm 1.54}$	$69.70_{\pm 0.96}$	$63.74_{\pm 1.22}$	$64.23_{\pm 0.80}$
MLS	Near	Baseline	$70.60_{\pm 0.96}$	$60.99_{\pm 0.66}$	$57.88_{\pm 0.41}$	$71.34_{\pm 0.56}$	$59.70_{\pm 0.65}$	$57.27_{\pm 1.08}$
		ours	$70.77_{\pm 0.90}$	$61.24_{\pm 0.61}$	$58.31_{\pm 0.38}$	$71.32_{\pm 0.58}$	$59.70_{\pm 0.68}$	$57.36_{\pm 1.10}$
	Far	Baseline	$70.03_{\pm 1.34}$	$60.94_{\pm 0.72}$	$57.87_{\pm 0.41}$	$70.95_{\pm 0.54}$	$59.50_{\pm 0.64}$	$57.14_{\pm 1.05}$
		ours	$70.23_{\pm 1.32}$	$61.18_{\pm 0.68}$	$58.29_{\pm 0.38}$	$70.93_{\pm 0.56}$	$59.50_{\pm 0.69}$	$57.23_{\pm 1.06}$
ReAct	Near	Baseline	$70.80_{\pm 0.33}$	$58.67_{\pm 1.51}$	$56.65_{\pm 0.98}$	$70.79_{\pm 0.20}$	$59.60_{\pm 0.86}$	$56.43_{\pm 0.56}$
		ours	$71.06_{\pm 0.32}$	$59.23_{\pm 1.56}$	$57.09_{\pm 1.14}$	$70.76_{\pm 0.20}$	$59.57_{\pm 0.86}$	$56.48_{\pm 0.62}$
	Far	Baseline	$71.06_{\pm 0.14}$	$58.83_{\pm 1.50}$	$56.64_{\pm 0.98}$	$70.90_{\pm 0.18}$	$59.78_{\pm 0.78}$	$56.46_{\pm 0.54}$
		ours	$71.33_{\pm 0.02}$	$59.41_{\pm 1.55}$	$57.08_{\pm 1.14}$	$70.87_{\pm 0.15}$	$59.72_{\pm 0.81}$	$56.51_{\pm 0.63}$
KNN	Near	Baseline	$71.88_{\pm 0.22}$	$61.32_{\pm 1.64}$	$56.66_{\pm 2.80}$	$71.72_{\pm 0.13}$	$58.73_{\pm 0.84}$	$57.32_{\pm 0.82}$
		ours	$72.08_{\pm 0.26}$	$61.59_{\pm 1.62}$	$57.06_{\pm 2.75}$	$71.72_{\pm 0.10}$	$58.74_{\pm 0.81}$	$57.35_{\pm 0.80}$
	Far	Baseline	$73.06_{\pm 0.05}$	$61.55_{\pm 1.65}$	$56.71_{\pm 2.78}$	$72.95_{\pm 0.47}$	$58.93_{\pm 0.89}$	$57.27_{\pm 0.82}$
		ours	$73.24_{\pm 0.09}$	$61.80_{\pm 1.64}$	$57.11_{\pm 2.72}$	$72.94_{\pm 0.49}$	$58.93_{\pm 0.85}$	$57.30_{\pm 0.80}$
Ash	Near	Baseline	$68.44_{\pm 0.66}$	$57.60_{\pm 0.15}$	$55.48_{\pm 0.26}$	$67.87_{\pm 0.16}$	$57.02_{\pm 0.88}$	$56.17_{\pm 0.91}$
		ours	$68.58_{\pm 0.68}$	$57.83_{\pm 0.08}$	$55.73_{\pm 0.20}$	$67.85_{\pm 0.15}$	$57.08_{\pm 0.95}$	$56.21_{\pm 0.85}$
	Far	Baseline	$70.54_{\pm 0.15}$	$57.90_{\pm 0.04}$	$55.52_{\pm 0.26}$	$70.21_{\pm 0.55}$	$57.30_{\pm 0.68}$	$56.23_{\pm 0.92}$
		ours	$70.70_{\pm 0.18}$	$58.17_{\pm 0.07}$	$55.77_{\pm 0.19}$	$70.17_{\pm 0.51}$	$57.37_{\pm 0.74}$	$56.27_{\pm 0.86}$

Table 29. **Top1 Accuracy (↑) of our OSLS correction model on the CIFAR100 dataset with Near OOD datasets and Far OOD datasets comparison under Dirichlet ID and OOD label shift. Settings in which our model outperforms baselines are colored in gray.** Our model outperforms baselines under most label shift settings. Each metric is averaged among the corresponding OOD test set (Tab. 1) and over three independent ID classifiers.