On Multi-Domain Long-Tailed Recognition, Imbalanced Domain Generalization and Beyond

Yuzhe Yang
MIT

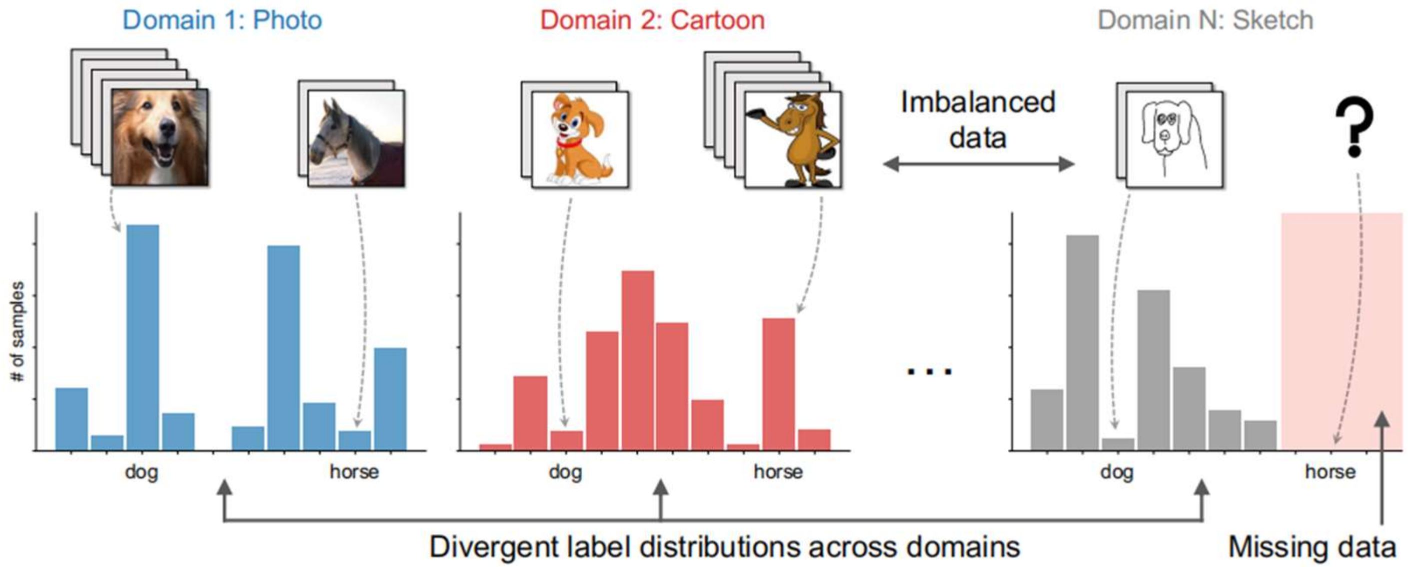
Hao Wang
Rutgers University

Dina Katabi
MIT

ECCV2022

Introduce

Multi-Domain Long-Tailed Recognition (MDLT): learning from multi-domain imbalanced data, with each domain having its own imbalanced label distribution, and generalizing to a test set that is balanced over all domain-class pairs.



- the label distribution for each domain is likely different from other domains
- multi-domain data inherently involves domain shift
- MDLT motivates zero-shot generalization within and across domains

Domain-Class Transferability Graph

label space: $\mathcal{C} = \{1, \dots, C\}$

domain space: $\mathcal{D} = \{1, \dots, D\}$

training set: $\mathcal{S} = \{(\mathbf{x}_i, c_i, d_i)\}_{i=1}^N$

domain-class pair : (d, c)

the set of all domain-class pairs:

$$\mathcal{M} = \mathcal{D} \times \mathcal{C} := \{(d, c) : d \in \mathcal{D}, c \in \mathcal{C}\}$$

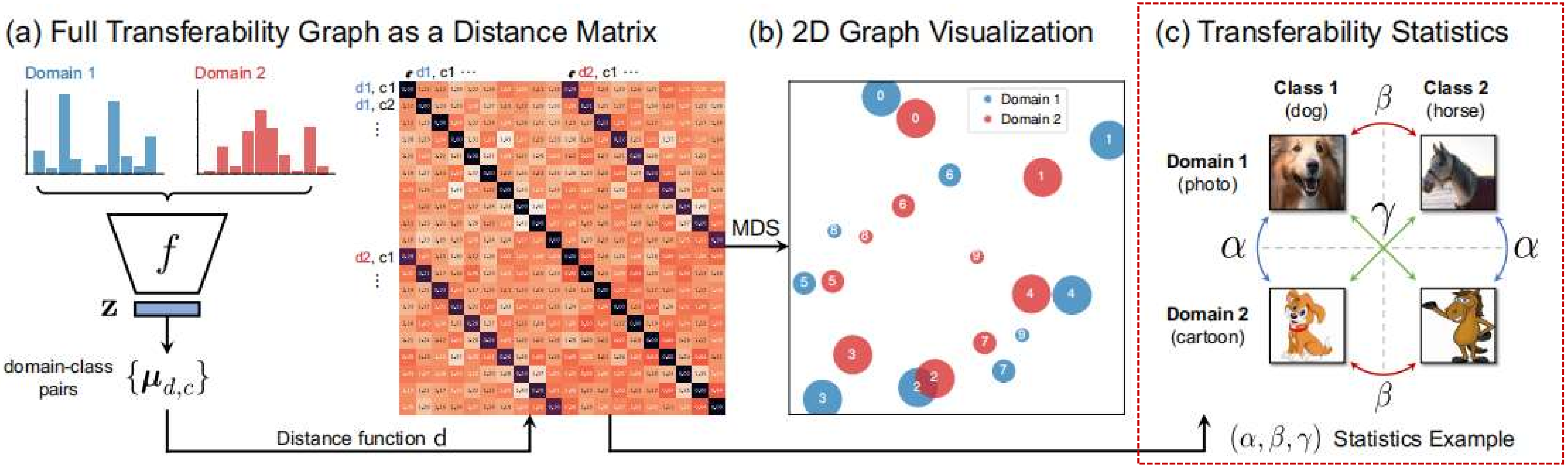
Definition 1 (Transferability). *Given a learned model and a distance function* $\mathsf{d} : \mathbb{R}^h \times \mathbb{R}^h \to \mathbb{R}$ *in the feature space, the transferability from domain-class pair* (d, c) *to* (d', c') *is:*

$$\operatorname{trans}\big((d, c), (d', c')\big) \triangleq \mathbb{E}_{\mathbf{z} \in \mathcal{Z}_{d,c}}\left[\mathsf{d}\left(\mathbf{z}, \boldsymbol{\mu}_{d',c'}\right)\right],$$

where $\boldsymbol{\mu}_{d',c'} \triangleq \mathbb{E}_{\mathbf{z}' \in \mathcal{Z}_{d',c'}}[\mathbf{z}']$ *is the first order statistics (i.e., mean) of* (d', c').

Definition 2 (Transferability Graph). *The transferability graph for a learned model is defined as* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, *where the vertices,* $\mathcal{V} \subseteq \{\boldsymbol{\mu}_{d,c}\}$, *represents the domain-class pairs, and the edges,* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, *are assigned weights equal to* $\operatorname{trans}\big((d, c), (d', c')\big)$.

Domain-Class Transferability Graph



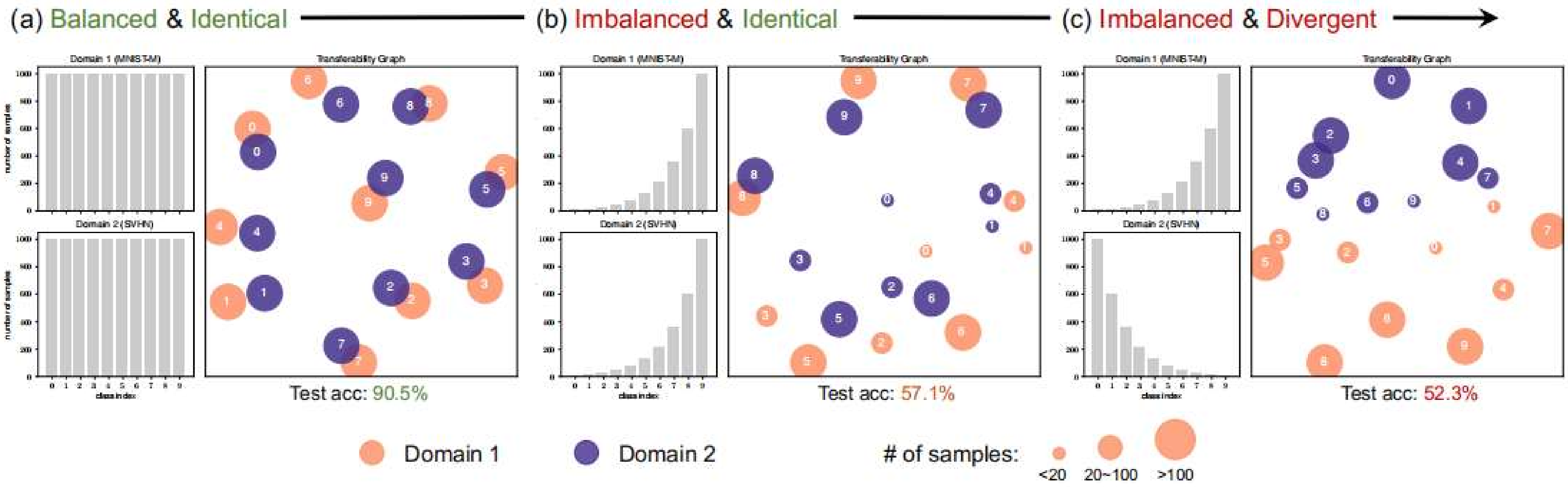
Definition 3 ((α, β, γ) Transferability Statistics). *The transferability graph can be summarized by the following transferability statistics:*

$$\textit{Different domains, same class:} \quad \alpha = \mathbb{E}_c \mathbb{E}_d \mathbb{E}_{d' \neq d} \left[\text{trans}\big((d,c),(d',c)\big) \right].$$

$$\textit{Same domain, different classes:} \quad \beta = \mathbb{E}_d \mathbb{E}_c \mathbb{E}_{c' \neq c} \left[\text{trans}\big((d,c),(d,c')\big) \right].$$

$$\textit{Different domains, different classes:} \quad \gamma = \mathbb{E}_d \mathbb{E}_{d' \neq d} \mathbb{E}_c \mathbb{E}_{c' \neq c} \left[\text{trans}\big((d,c),(d',c')\big) \right].$$

Transferability Statistics

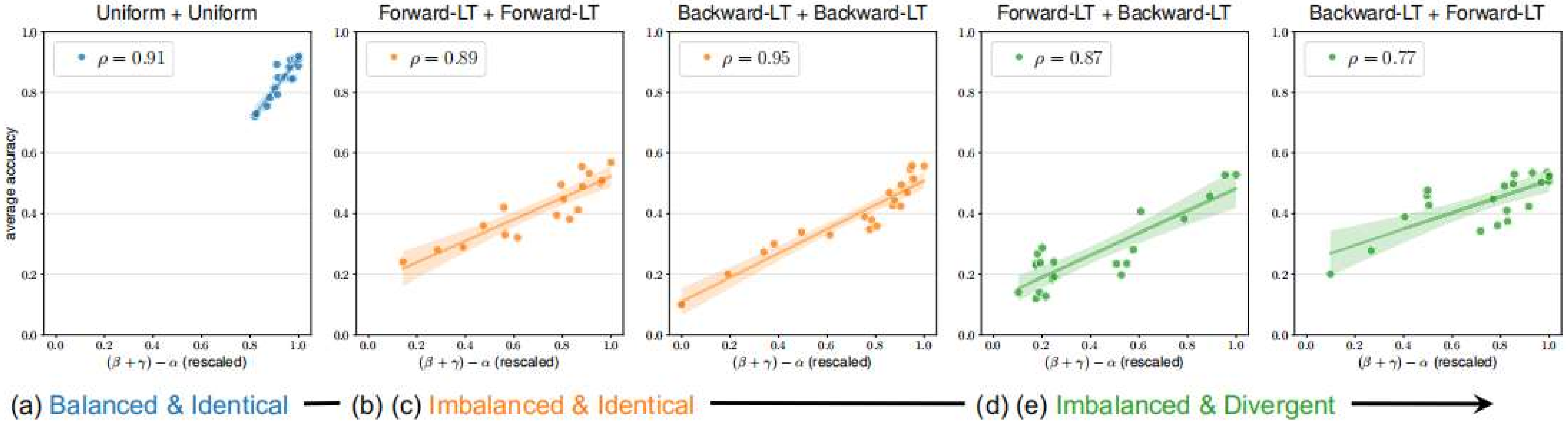


- When the per-domain label distributions are balanced and identical across domains, it does not prohibit the model from learning discriminative features of high accuracy.
- If the label distributions are imbalanced but identical, ERM is still able to align similar classes in the two domains.
- When the labels are both imbalanced and mismatched across domains, the learned features are no longer transferable.

$$\rightarrow (\alpha, \beta, \gamma)$$

Transferability Statistics

train 20 ERM models with varying hyperparameters



The (α , β , γ) statistics can characterize a model's performance in MDLT

Data imbalance increases the risk of learning less transferable features

BoDA

Domain-Class Distribution Alignment Loss:

positive cross-domain pairs (α)

$$\mathcal{L}_{\texttt{DA}}(\mathcal{Z},\{\boldsymbol{\mu}\}) = \sum_{\mathbf{z}_i\in\mathcal{Z}} \frac{-1}{|\mathcal{D}|-1} \sum_{d\in\mathcal{D}\setminus\{d_i\}} \log \frac{\boxed{\exp\left(-\mathrm{d}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c_i})\right)}}{\boxed{\sum_{(d',c')\in\mathcal{M}\setminus\{(d_i,c_i)\}} \exp\left(-\mathrm{d}(\mathbf{z}_i,\boldsymbol{\mu}_{d',c'})\right)}}$$

negative cross-class pairs (β,γ)

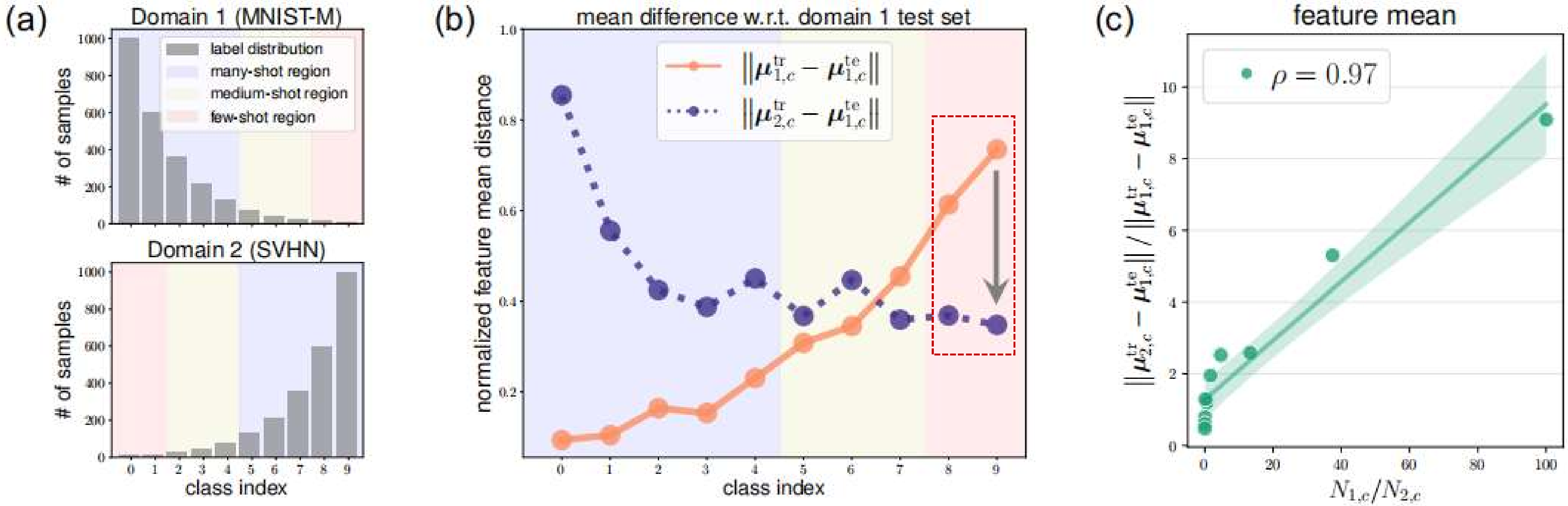
Balanced Domain-Class Distribution Alignment (BoDA):

$$\mathcal{L}_{\texttt{BoDA}}(\mathcal{Z},\{\boldsymbol{\mu}\}) = \sum_{\mathbf{z}_i\in\mathcal{Z}} \frac{-1}{|\mathcal{D}|-1} \sum_{d\in\mathcal{D}\setminus\{d_i\}} \log \frac{\exp\left(-\widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c_i})\right)}{\sum_{(d',c')\in\mathcal{M}\setminus\{(d_i,c_i)\}} \exp\left(-\widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d',c'})\right)}, \quad \widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c}) = \frac{\mathrm{d}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c})}{N_{d_i,c_i}}$$

Theorem 1 ($\mathcal{L}_{\texttt{BoDA}}$ as an Upper Bound). *Given a multi-domain long-tailed dataset $\mathcal{S}$ with domain label space $\mathcal{D}$ and class label space $\mathcal{C}$ satisfying $|\mathcal{D}|>1$ and $|\mathcal{C}|>1$, let $\mathcal{Z}$ be the representation set of all training samples, and (α,β,γ) be the transferability statistics for $\mathcal{S}$ defined in Definition 3. It holds that*

$$\mathcal{L}_{\texttt{BoDA}}(\mathcal{Z},\{\boldsymbol{\mu}\}) \geq N\log\left(|\mathcal{D}|-1+|\mathcal{D}|(|\mathcal{C}|-1)\exp\left(\frac{|\mathcal{C}||\mathcal{D}|}{N}\cdot\alpha-\frac{|\mathcal{C}|}{N}\cdot\beta-\frac{|\mathcal{C}|(|\mathcal{D}|-1)}{N}\cdot\gamma\right)\right). \tag{3}$$

BoDA



$$\mathcal{L}_{\texttt{BoDA}}(\mathcal{Z},\{\boldsymbol{\mu}\}) = \sum_{\mathbf{z}_i\in\mathcal{Z}} \frac{-1}{|\mathcal{D}|-1} \sum_{d\in\mathcal{D}\setminus\{d_i\}} \log \frac{\exp\left(-\widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c_i})\right)}{\sum_{(d',c')\in\mathcal{M}\setminus\{(d_i,c_i)\}}\exp\left(-\widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d',c'})\right)}, \quad \widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c}) = \frac{\mathrm{d}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c})}{N_{d_i,c_i}}$$

$$\widetilde{\mathcal{L}}_{\texttt{BoDA}}(\mathcal{Z},\{\boldsymbol{\mu}\}) = \sum_{\mathbf{z}_i\in\mathcal{Z}} \frac{-1}{|\mathcal{D}|-1} \sum_{d\in\mathcal{D}\setminus\{d_i\}} \log \frac{\exp\left(-\lambda^{d,c_i}_{d_i,c_i}\,\widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d,c_i})\right)}{\sum_{(d',c')\in\mathcal{M}\setminus\{(d_i,c_i)\}}\exp\left(-\lambda^{d',c'}_{d_i,c_i}\,\widetilde{\mathrm{d}}(\mathbf{z}_i,\boldsymbol{\mu}_{d',c'})\right)}, \quad \boxed{\lambda^{d',c'}_{d,c} = \left(\frac{N_{d',c'}}{N_{d,c}}\right)^{\nu}}$$

BoDA

Algorithm 1 Balanced Domain-Class Distribution Alignment (BoDA)

```
Input: Training set D = {(x_i, c_i, d_i)}_{i=1}^N, all domain-class pairs M = {(d,c)}, encoder f, classifier
g, total training epochs E, calibration parameter ν, loss weight ω, momentum α
for all (d,c) ∈ M do
  Initialize the feature statistics {μ_{d,c}^{(0)}, Σ_{d,c}^{(0)}}
end for
for e = 0 to E do
  repeat
    Sample a mini-batch {(x_i, c_i, d_i)}_{i=1}^m from D
    for i = 1 to m (in parallel) do
      z_i = f(x_i)
      ĉ_i = g(z_i)
    end for
    Calculate L̃_BoDA using {z_i} based on Eqn. (4)
    Calculate L_CE using (1/m) Σ_{i=1}^m L(ĉ_i, c_i)
    Do one training step with loss L_CE + ω L̃_BoDA
  until iterate over all training samples at current epoch e
  /* Update feature statistics with momentum updating */
  for all (d,c) ∈ M do
    Estimate current feature statistics {μ_{d,c}, Σ_{d,c}}
    μ_{d,c}^{(e+1)} ← α × μ_{d,c}^{(e)} + (1 − α) × μ_{d,c}
    Σ_{d,c}^{(e+1)} ← α × Σ_{d,c}^{(e)} + (1 − α) × Σ_{d,c}
  end for
end for
```

$\widetilde{\mathcal{L}}_{\texttt{BoDA}}$ $\quad \mathrm{d}(\mathbf{z}, \boldsymbol{\mu}_{d,c}) = \sqrt{(\mathbf{z}-\boldsymbol{\mu}_{d,c})^\top(\mathbf{z}-\boldsymbol{\mu}_{d,c})}$

$\widetilde{\mathcal{L}}_{\texttt{BoDA-M}}$ $\quad \mathrm{d}(\mathbf{z}, \{\boldsymbol{\mu}_{d,c}, \boldsymbol{\Sigma}_{d,c}\}) = \sqrt{(\mathbf{z}-\boldsymbol{\mu}_{d,c})^\top \boldsymbol{\Sigma}_{d,c}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{d,c})}$

$\texttt{BoDA}_r$

couple representation and classifier learning

$\texttt{BoDA}_{r,c}$

decouple representation from classifier learning

Experiments

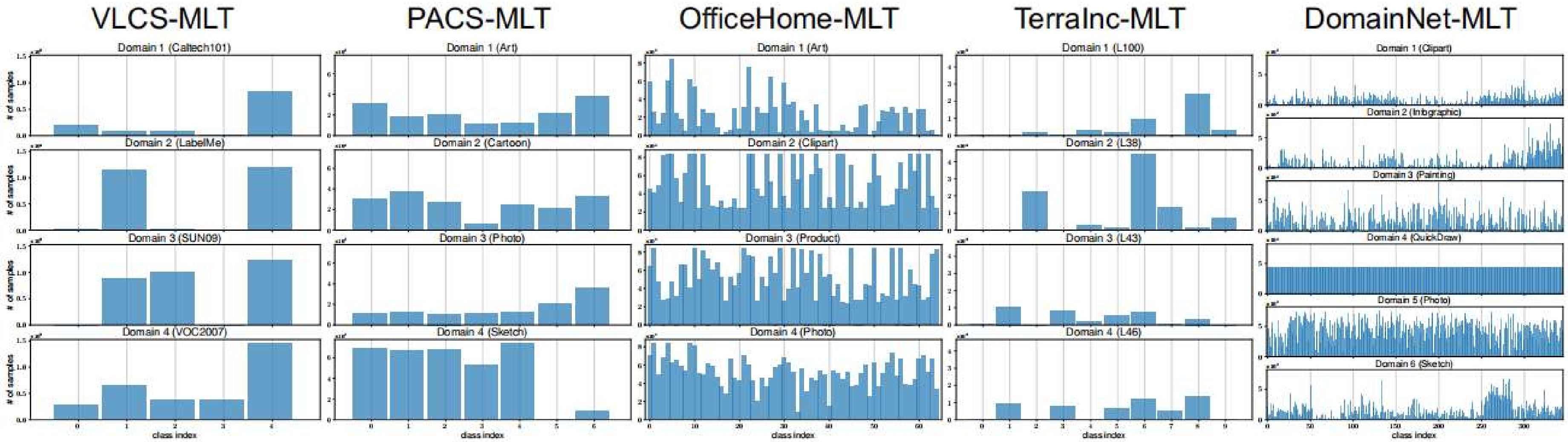


Figure 6: Overview of training set label distribution for five MDLT datasets. We set up MDLT benchmarks from datasets traditionally used for DG, and make validation/test sets balanced across all domain-class pairs. More details are provided in Appendix D.

Experiments

Table 2: Results on VLCS-MLT.

Algorithm	Accuracy (by domain) Average	Accuracy (by domain) Worst	Accuracy (by shot) Many	Accuracy (by shot) Medium	Accuracy (by shot) Few	Accuracy (by shot) Zero
ERM [46]	76.3 ±0.4	53.6 ±1.1	84.6 ±0.5	76.6 ±0.4	–	32.9 ±0.4
IRM [1]	76.5 ±0.2	52.3 ±0.7	85.3 ±0.6	75.5 ±1.0	–	33.5 ±1.0
GroupDRO [40]	76.7 ±0.4	54.1 ±1.3	85.3 ±0.9	76.2 ±1.0	–	34.5 ±2.0
Mixup [50]	75.9 ±0.1	52.7 ±1.3	84.4 ±0.2	77.1 ±0.6	–	29.2 ±1.4
MLDG [28]	76.9 ±0.2	53.6 ±0.5	84.9 ±0.3	77.5 ±1.0	–	34.4 ±0.9
CORAL [45]	75.9 ±0.5	51.6 ±0.7	84.3 ±0.6	75.5 ±0.5	–	34.5 ±0.8
MMD [29]	76.3 ±0.6	53.4 ±0.3	84.5 ±0.8	77.1 ±0.5	–	32.7 ±0.3
DANN [15]	77.5 ±0.1	54.1 ±0.3	85.9 ±0.5	76.0 ±0.4	–	38.0 ±2.3
CDANN [31]	76.6 ±0.4	53.6 ±0.4	84.4 ±0.7	77.3 ±0.8	–	35.0 ±0.8
MTL [4]	76.3 ±0.3	52.9 ±0.5	84.8 ±0.9	76.2 ±0.6	–	33.3 ±1.4
SagNet [35]	76.3 ±0.2	52.3 ±0.2	85.3 ±0.3	75.1 ±0.2	–	32.9 ±0.3
Fish [42]	77.5 ±0.3	54.3 ±0.4	86.2 ±0.5	76.0 ±0.4	–	35.6 ±2.2
Focal [32]	75.6 ±0.4	52.3 ±0.2	84.0 ±0.2	75.5 ±0.6	–	32.7 ±0.9
CBLoss [10]	76.8 ±0.3	52.5 ±0.5	84.8 ±0.7	77.5 ±1.4	–	33.2 ±1.6
LDAM [6]	77.5 ±0.1	52.9 ±0.2	**86.5** ±0.4	75.5 ±0.5	–	35.2 ±0.6
BSoftmax [39]	76.7 ±0.5	52.9 ±0.9	84.4 ±0.9	78.2 ±0.6	–	34.3 ±0.9
SSP [52]	76.1 ±0.3	52.3 ±1.0	83.8 ±0.3	76.0 ±1.2	–	37.1 ±0.7
CRT [23]	76.3 ±0.2	51.4 ±0.3	84.5 ±0.1	77.3 ±0.0	–	31.7 ±1.0
BoDA$_r$	76.9 ±0.5	51.4 ±0.3	85.3 ±0.3	77.3 ±0.2	–	33.3 ±0.5
BoDA-M$_r$	77.5 ±0.3	53.4 ±0.3	85.8 ±0.2	77.3 ±0.2	–	35.7 ±0.7
BoDA$_{r,c}$	77.3 ±0.2	53.4 ±0.3	85.3 ±0.3	78.0 ±0.2	–	38.6 ±0.7
BoDA-M$_{r,c}$	**78.2** ±0.4	**55.4** ±0.5	85.3 ±0.3	**79.3** ±0.6	–	**43.3** ±1.1
BoDA *vs.* ERM	+1.9	+1.8	+0.7	+2.7	–	+10.4

Table 3: Results on PACS-MLT.

Algorithm	Accuracy (by domain) Average	Accuracy (by domain) Worst	Accuracy (by shot) Many	Accuracy (by shot) Medium	Accuracy (by shot) Few	Accuracy (by shot) Zero
ERM [46]	97.1 ±0.1	95.8 ±0.2	97.1 ±0.0	97.0 ±0.0	98.0 ±0.9	–
IRM [1]	96.7 ±0.2	95.2 ±0.4	96.8 ±0.2	96.7 ±0.7	94.7 ±1.4	–
GroupDRO [40]	97.0 ±0.1	95.3 ±0.4	97.3 ±0.1	95.3 ±1.2	94.7 ±3.6	–
Mixup [50]	96.7 ±0.2	95.1 ±0.2	97.0 ±0.1	96.7 ±0.3	91.3 ±2.7	–
MLDG [28]	96.6 ±0.1	94.1 ±0.3	96.8 ±0.1	96.3 ±0.7	92.7 ±0.5	–
CORAL [45]	96.6 ±0.5	94.3 ±0.7	96.6 ±0.5	97.0 ±0.8	94.7 ±0.5	–
MMD [29]	96.9 ±0.1	96.2 ±0.2	96.9 ±0.2	97.0 ±0.0	96.7 ±0.5	–
DANN [15]	96.5 ±0.0	94.3 ±0.1	96.5 ±0.1	98.0 ±0.0	94.7 ±2.4	–
CDANN [31]	96.1 ±0.1	94.5 ±0.2	96.1 ±0.1	96.3 ±0.5	94.0 ±0.9	–
MTL [4]	96.7 ±0.2	94.5 ±0.6	96.8 ±0.1	95.3 ±1.7	97.3 ±1.1	–
SagNet [35]	**97.2** ±0.1	95.2 ±0.3	**97.4** ±0.1	96.7 ±0.5	95.3 ±0.5	–
Fish [42]	96.9 ±0.2	95.2 ±0.2	97.0 ±0.1	97.0 ±0.5	94.7 ±1.1	–
Focal [32]	96.5 ±0.2	94.6 ±0.7	96.6 ±0.1	95.0 ±1.7	96.7 ±0.5	–
CBLoss [10]	96.9 ±0.1	95.1 ±0.4	96.8 ±0.2	97.0 ±1.2	**100.0** ±0.0	–
LDAM [6]	96.5 ±0.2	94.7 ±0.2	96.6 ±0.1	95.7 ±1.4	96.0 ±0.0	–
BSoftmax [39]	96.9 ±0.3	95.6 ±0.3	96.6 ±0.4	**98.7** ±0.7	99.3 ±0.5	–
SSP [52]	96.9 ±0.2	95.4 ±0.4	96.7 ±0.2	98.3 ±0.5	98.0 ±0.9	–
CRT [23]	96.3 ±0.1	94.9 ±0.1	96.3 ±0.1	97.3 ±0.3	94.0 ±0.9	–
BoDA$_r$	97.0 ±0.1	95.1 ±0.4	97.0 ±0.1	96.3 ±0.5	98.0 ±0.9	–
BoDA-M$_r$	97.1 ±0.1	94.9 ±0.1	97.3 ±0.1	96.3 ±0.5	96.0 ±0.0	–
BoDA$_{r,c}$	**97.2** ±0.1	95.7 ±0.3	**97.4** ±0.1	97.0 ±0.0	94.7 ±1.1	–
BoDA-M$_{r,c}$	97.1 ±0.2	**96.3** ±0.1	97.1 ±0.0	97.0 ±0.8	96.0 ±0.0	–
BoDA *vs.* ERM	+0.1	+0.5	+0.3	+0.0	-2.0	–

Experiments

Table 4: Results on `OfficeHome-MLT`.

Algorithm	Accuracy (by domain)		Accuracy (by shot)			
	Average	Worst	Many	Medium	Few	Zero
ERM [46]	80.7 ±0.0	71.3 ±0.1	87.8 ±0.2	81.0 ±0.2	63.1 ±0.1	63.3 ±7.2
IRM [1]	80.6 ±0.4	70.7 ±0.2	87.6 ±0.4	81.5 ±0.4	61.1 ±0.9	56.7 ±1.4
GroupDRO [40]	80.1 ±0.3	68.7 ±0.9	88.1 ±0.2	80.8 ±0.4	59.8 ±1.2	51.7 ±3.6
Mixup [50]	81.2 ±0.2	72.3 ±0.6	87.9 ±0.4	81.8 ±0.1	64.1 ±0.4	60.0 ±4.1
MLDG [28]	80.4 ±0.2	70.2 ±0.6	87.1 ±0.1	81.3 ±0.3	61.3 ±1.0	61.7 ±1.4
CORAL [45]	81.9 ±0.1	**72.7** ±0.6	87.9 ±0.1	83.0 ±0.1	63.5 ±0.7	65.0 ±2.4
MMD [29]	78.4 ±0.4	67.7 ±0.8	85.2 ±0.2	79.4 ±0.7	58.8 ±0.4	56.7 ±3.6
DANN [15]	79.2 ±0.2	70.2 ±0.9	86.2 ±0.1	80.0 ±0.1	60.3 ±1.1	61.7 ±5.9
CDANN [31]	79.0 ±0.2	69.4 ±0.3	86.4 ±0.6	79.8 ±0.1	58.9 ±0.8	50.0 ±4.7
MTL [4]	79.5 ±0.2	69.8 ±0.6	87.3 ±0.3	79.8 ±0.2	61.1 ±0.2	51.7 ±2.7
SagNet [35]	80.9 ±0.1	70.5 ±0.5	87.8 ±0.4	81.9 ±0.1	61.2 ±0.9	56.7 ±3.6
Fish [42]	81.3 ±0.3	71.3 ±0.7	**88.2** ±0.2	81.9 ±0.3	63.2 ±0.8	61.7 ±1.4
Focal [32]	77.9 ±0.0	67.6 ±0.4	86.5 ±0.3	78.3 ±0.1	57.4 ±0.3	46.7 ±3.6
CBLoss [10]	79.8 ±0.2	69.5 ±0.7	86.6 ±0.4	80.6 ±0.2	61.1 ±1.4	65.0 ±2.4
LDAM [6]	80.3 ±0.2	69.9 ±0.5	87.1 ±0.2	81.3 ±0.3	61.1 ±0.2	51.7 ±2.7
BSoftmax [39]	80.4 ±0.2	70.9 ±0.5	86.7 ±0.5	81.3 ±0.3	62.4 ±1.0	60.0 ±4.1
SSP [52]	81.1 ±0.3	71.1 ±0.3	87.3 ±0.6	82.3 ±0.3	61.6 ±0.7	63.3 ±1.4
CRT [23]	81.2 ±0.0	72.5 ±0.2	87.7 ±0.1	81.8 ±0.1	64.0 ±0.1	65.0 ±2.4
BoDA$_r$	81.5 ±0.1	71.8 ±0.1	87.7 ±0.2	82.3 ±0.1	**64.2** ±0.3	63.3 ±1.4
BoDA-M$_r$	81.9 ±0.2	71.6 ±0.2	87.3 ±0.3	83.4 ±0.2	62.3 ±0.3	65.0 ±2.4
BoDA$_{r,c}$	82.3 ±0.1	72.3 ±0.3	87.1 ±0.2	**83.9** ±0.3	63.2 ±0.2	65.0 ±2.4
BoDA-M$_{r,c}$	**82.4** ±0.2	72.3 ±0.3	87.7 ±0.1	**83.9** ±0.6	**64.2** ±0.3	**66.7** ±2.7
BoDA *vs.* ERM	+1.7	+1.0	-0.1	+2.9	+1.1	+3.4

Table 5: Results on `TerraInc-MLT`.

Algorithm	Accuracy (by domain)		Accuracy (by shot)			
	Average	Worst	Many	Medium	Few	Zero
ERM [46]	75.3 ±0.3	67.4 ±0.3	85.6 ±0.8	69.6 ±3.2	66.1 ±2.4	14.4 ±2.8
IRM [1]	73.3 ±0.7	64.3 ±1.3	83.5 ±0.6	70.0 ±1.8	58.3 ±3.4	20.1 ±1.4
GroupDRO [40]	72.0 ±0.4	66.6 ±0.2	84.7 ±1.1	64.6 ±4.7	38.9 ±1.2	13.5 ±1.1
Mixup [50]	71.1 ±0.7	60.4 ±1.1	83.2 ±0.7	60.0 ±0.6	56.1 ±3.0	12.2 ±2.1
MLDG [28]	76.6 ±0.2	66.9 ±0.5	86.1 ±0.6	73.8 ±3.9	70.6 ±3.7	18.8 ±2.4
CORAL [45]	76.4 ±0.5	67.8 ±0.9	86.3 ±0.3	77.5 ±3.1	66.1 ±2.0	11.0 ±1.4
MMD [29]	73.3 ±0.4	63.7 ±1.1	84.0 ±0.4	67.9 ±2.7	60.6 ±1.6	13.6 ±2.6
DANN [15]	68.7 ±0.9	61.1 ±1.0	79.6 ±1.2	62.5 ±8.1	48.9 ±2.8	13.3 ±1.1
CDANN [31]	70.3 ±0.5	63.9 ±1.0	83.5 ±0.8	50.0 ±4.2	43.9 ±4.7	20.4 ±3.1
MTL [4]	75.0 ±0.7	67.7 ±1.4	85.2 ±0.7	73.8 ±1.6	61.1 ±2.8	12.4 ±4.0
SagNet [35]	75.1 ±1.6	66.5 ±2.1	85.5 ±0.9	77.1 ±5.0	57.8 ±4.3	13.0 ±3.4
Fish [42]	75.3 ±0.5	66.3 ±0.5	85.8 ±0.2	73.3 ±3.9	61.1 ±3.0	13.7 ±3.3
Focal [32]	75.7 ±0.4	65.3 ±1.1	85.7 ±0.3	76.2 ±3.9	68.9 ±3.2	12.6 ±1.9
CBLoss [10]	78.0 ±0.4	68.3 ±2.0	85.0 ±0.1	89.2 ±1.2	83.9 ±2.5	9.3 ±3.9
LDAM [6]	74.7 ±0.9	64.1 ±1.4	85.1 ±0.6	70.8 ±3.5	67.8 ±1.2	11.1 ±2.4
BSoftmax [39]	76.7 ±1.0	65.6 ±1.3	83.4 ±0.8	90.8 ±0.9	78.3 ±3.9	12.6 ±2.4
SSP [52]	78.5 ±0.7	67.3 ±0.4	85.5 ±1.0	87.8 ±0.9	82.6 ±1.2	13.2 ±2.8
CRT [23]	81.6 ±0.1	70.0 ±0.4	**89.7** ±0.2	90.4 ±0.3	83.9 ±0.5	12.9 ±0.0
BoDA$_r$	78.6 ±0.4	68.5 ±0.3	86.4 ±0.1	85.0 ±1.0	80.0 ±0.9	13.7 ±2.1
BoDA-M$_r$	79.4 ±0.6	71.3 ±0.4	88.4 ±0.3	76.2 ±2.7	88.3 ±1.6	14.4 ±1.4
BoDA$_{r,c}$	82.3 ±0.3	68.5 ±0.6	89.2 ±0.2	**92.5** ±0.9	88.3 ±1.2	21.3 ±0.7
BoDA-M$_{r,c}$	**83.0** ±0.4	**74.6** ±0.7	89.2 ±0.2	91.2 ±0.6	**91.7** ±2.0	**21.7** ±1.4
BoDA *vs.* ERM	+7.7	+7.2	+3.6	+22.9	+25.6	+7.3

Experiments

Table 6: Results on `DomainNet-MLT`.

Algorithm	Accuracy (by domain)		Accuracy (by shot)			
	Average	Worst	Many	Medium	Few	Zero
ERM [46]	58.6 ±0.2	29.4 ±0.3	66.0 ±0.1	56.1 ±0.1	35.9 ±0.5	27.6 ±0.3
IRM [1]	57.1 ±0.1	27.6 ±0.1	64.7 ±0.1	54.3 ±0.3	33.5 ±0.3	25.8 ±0.3
GroupDRO [40]	53.6 ±0.1	25.9 ±0.2	61.8 ±0.1	49.1 ±0.3	30.7 ±0.7	22.0 ±0.1
Mixup [50]	57.6 ±0.1	28.7 ±0.0	64.9 ±0.2	54.5 ±0.1	35.6 ±0.2	27.3 ±0.3
MLDG [28]	58.5 ±0.0	28.7 ±0.1	66.0 ±0.1	55.7 ±0.1	35.3 ±0.2	26.9 ±0.3
CORAL [45]	59.4 ±0.1	30.1 ±0.4	66.4 ±0.1	57.1 ±0.0	37.7 ±0.6	29.9 ±0.2
MMD [29]	56.7 ±0.0	27.2 ±0.2	64.2 ±0.1	54.0 ±0.0	33.9 ±0.2	25.4 ±0.2
DANN [15]	55.8 ±0.1	26.9 ±0.4	63.0 ±0.1	52.7 ±0.1	34.2 ±0.4	26.8 ±0.4
CDANN [31]	56.0 ±0.1	27.7 ±0.1	63.2 ±0.0	52.7 ±0.2	34.3 ±0.5	27.6 ±0.1
MTL [4]	58.6 ±0.1	29.3 ±0.2	65.9 ±0.1	56.0 ±0.4	35.4 ±0.1	28.2 ±0.3
SagNet [35]	58.9 ±0.0	29.4 ±0.2	66.3 ±0.1	56.4 ±0.0	36.2 ±0.3	27.2 ±0.4
Fish [42]	59.6 ±0.1	29.1 ±0.1	67.1 ±0.1	57.2 ±0.1	36.8 ±0.4	27.8 ±0.3
Focal [32]	57.8 ±0.2	27.5 ±0.1	65.2 ±0.2	55.1 ±0.2	35.8 ±0.1	26.3 ±0.1
CBLoss [10]	58.9 ±0.1	30.1 ±0.1	64.3 ±0.0	61.0 ±0.3	42.5 ±0.4	28.1 ±0.2
LDAM [6]	59.2 ±0.0	29.2 ±0.2	66.6 ±0.0	57.0 ±0.0	37.1 ±0.2	27.8 ±0.3
BSoftmax [39]	58.9 ±0.1	29.9 ±0.1	64.3 ±0.1	60.9 ±0.3	42.4 ±0.6	28.2 ±0.1
SSP [52]	59.7 ±0.0	31.6 ±0.2	64.3 ±0.1	62.6 ±0.1	45.0 ±0.3	30.5 ±0.0
CRT [23]	60.4 ±0.2	31.6 ±0.1	66.8 ±0.0	61.6 ±0.1	45.7 ±0.1	29.7 ±0.1
BoDA$_r$	60.1 ±0.2	32.6 ±0.1	65.7 ±0.2	60.6 ±0.1	42.6 ±0.3	30.5 ±0.2
BoDA-M$_r$	60.1 ±0.2	32.2 ±0.2	65.9 ±0.2	60.7 ±0.1	42.9 ±0.3	30.0 ±0.1
BoDA$_{r,c}$	**61.7** ±0.1	**33.4** ±0.1	**67.0** ±0.1	62.7 ±0.1	46.0 ±0.2	**32.2** ±0.3
BoDA-M$_{r,c}$	**61.7** ±0.2	33.3 ±0.1	**67.0** ±0.1	**63.0** ±0.3	**46.6** ±0.4	31.8 ±0.2
BoDA *vs.* ERM	+3.1	+4.0	+1.0	+6.9	+10.7	+4.6

Table 7: Results over all MDLT benchmarks.

Algorithm	`VLCS-MLT`	`PACS-MLT`	`OfficeHome-MLT`	`TerraInc-MLT`	`DomainNet-MLT`	Avg
ERM [46]	76.3 ±0.4	97.1 ±0.1	80.7 ±0.0	75.3 ±0.3	58.6 ±0.2	77.6
IRM [1]	76.5 ±0.2	96.7 ±0.2	80.6 ±0.4	73.3 ±0.7	57.1 ±0.1	76.8
GroupDRO [40]	76.7 ±0.4	97.0 ±0.1	80.1 ±0.3	72.0 ±0.4	53.6 ±0.1	75.9
Mixup [50]	75.9 ±0.1	96.7 ±0.2	81.2 ±0.2	71.1 ±0.7	57.6 ±0.1	76.5
MLDG [28]	76.9 ±0.2	96.6 ±0.1	80.4 ±0.2	76.6 ±0.2	58.5 ±0.0	77.8
CORAL [45]	75.9 ±0.5	96.6 ±0.5	81.9 ±0.1	76.4 ±0.5	59.4 ±0.1	78.0
MMD [29]	76.3 ±0.6	96.9 ±0.1	78.4 ±0.4	73.3 ±0.4	56.7 ±0.0	76.3
DANN [15]	77.5 ±0.1	96.5 ±0.0	79.2 ±0.2	68.7 ±0.9	55.8 ±0.1	75.5
CDANN [31]	76.6 ±0.4	96.1 ±0.1	79.0 ±0.2	70.3 ±0.5	56.0 ±0.1	75.6
MTL [4]	76.3 ±0.3	96.7 ±0.2	79.5 ±0.2	75.0 ±0.7	58.6 ±0.1	77.2
SagNet [35]	76.3 ±0.2	**97.2** ±0.1	80.9 ±0.1	75.1 ±1.6	58.9 ±0.0	77.7
Fish [42]	77.5 ±0.3	96.9 ±0.2	81.3 ±0.3	75.3 ±0.5	59.6 ±0.1	78.1
Focal [32]	75.6 ±0.4	96.5 ±0.2	77.9 ±0.0	75.7 ±0.4	57.8 ±0.2	76.7
CBLoss [10]	76.8 ±0.3	96.9 ±0.1	79.8 ±0.2	78.0 ±0.4	58.9 ±0.1	78.1
LDAM [6]	77.5 ±0.1	96.5 ±0.2	80.3 ±0.2	74.7 ±0.9	59.2 ±0.0	77.7
BSoftmax [39]	76.7 ±0.5	96.9 ±0.3	80.4 ±0.2	76.7 ±1.0	58.9 ±0.1	77.9
SSP [52]	76.1 ±0.3	96.9 ±0.2	81.1 ±0.3	78.5 ±0.7	59.7 ±0.0	78.5
CRT [23]	76.3 ±0.2	96.3 ±0.1	81.2 ±0.0	81.6 ±0.1	60.4 ±0.2	79.2
BoDA$_r$	76.9 ±0.5	97.0 ±0.1	81.5 ±0.1	78.6 ±0.4	60.1 ±0.2	78.8
BoDA-M$_r$	77.5 ±0.3	97.1 ±0.1	81.9 ±0.2	79.4 ±0.6	60.1 ±0.2	79.2
BoDA$_{r,c}$	77.3 ±0.2	**97.2** ±0.1	82.3 ±0.1	82.3 ±0.3	**61.7** ±0.1	80.2
BoDA-M$_{r,c}$	**78.2** ±0.4	97.1 ±0.2	**82.4** ±0.2	**83.0** ±0.4	**61.7** ±0.2	**80.5**
BoDA *vs.* ERM	+1.9	+0.1	+1.7	+7.7	+3.1	+2.9

Experiments

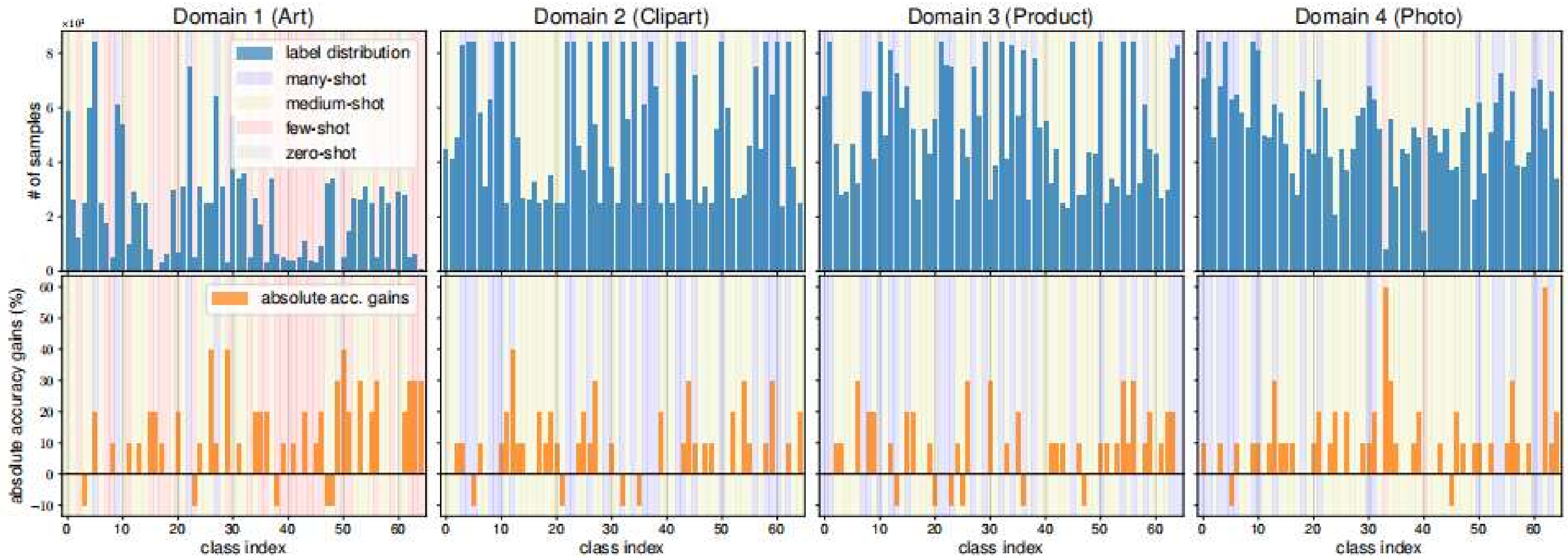


Figure 7: The absolute accuracy improvements of `BoDA` *vs.* ERM over all domain-class pairs on `OfficeHome-MLT`. `BoDA` establishes large improvements w.r.t. all regions, especially for the few-shot and zero-shot ones. Results for other datasets are in Appendix H.2.

Experiments

Table 24: Ablation study on effect of adding balanced distance in BoDA.

	VLCS-MLT	PACS-MLT	OfficeHome-MLT	TerraInc-MLT	DomainNet-MLT	**Avg**
DA	76.6 ±0.4	96.8 ±0.2	80.7 ±0.3	76.4 ±0.5	58.9 ±0.2	77.9
BoDA	**77.3** ±0.2	**97.2** ±0.1	**82.3** ±0.1	**82.3** ±0.3	**61.7** ±0.1	**80.2**
Gains	**+0.7**	**+0.4**	**+1.6**	**+5.9**	**+2.8**	**+2.3**

Table 25: Ablation study on effect of distance calibration coefficient $\lambda_{d,c}^{d',c'}$ in BoDA. We vary the value of ν and report the averaged results over all five MDLT datasets.

ν	0	0.5	0.7	0.9	1	1.1	1.2	1.5	ERM
BoDA	78.9	80.1	80.0	80.2	80.1	79.8	79.6	79.2	77.6

Experiments

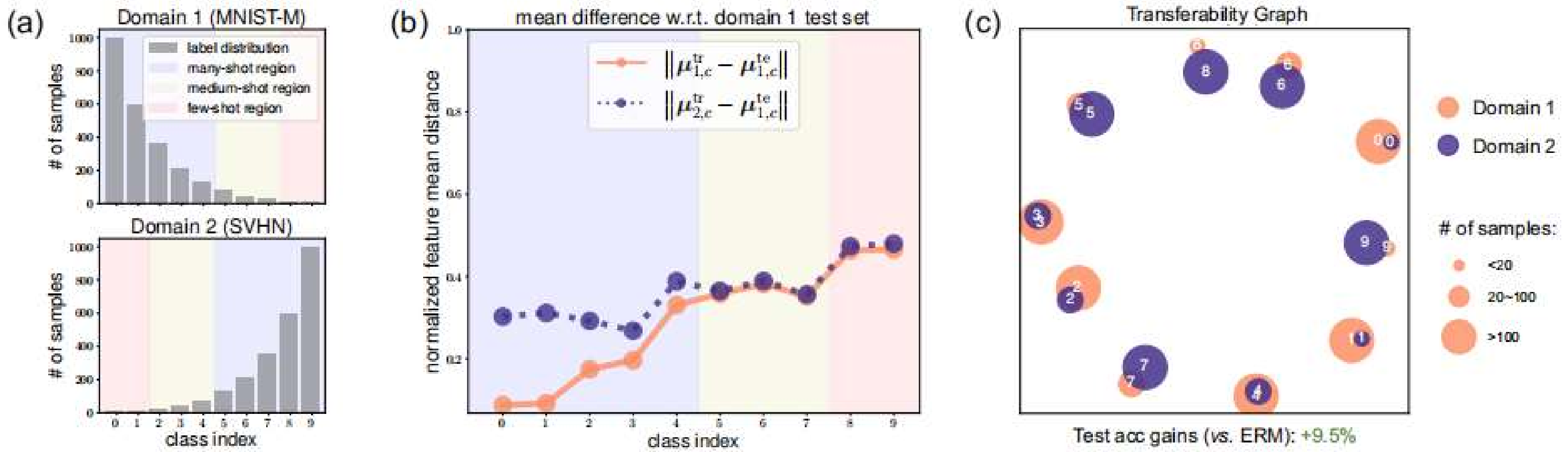


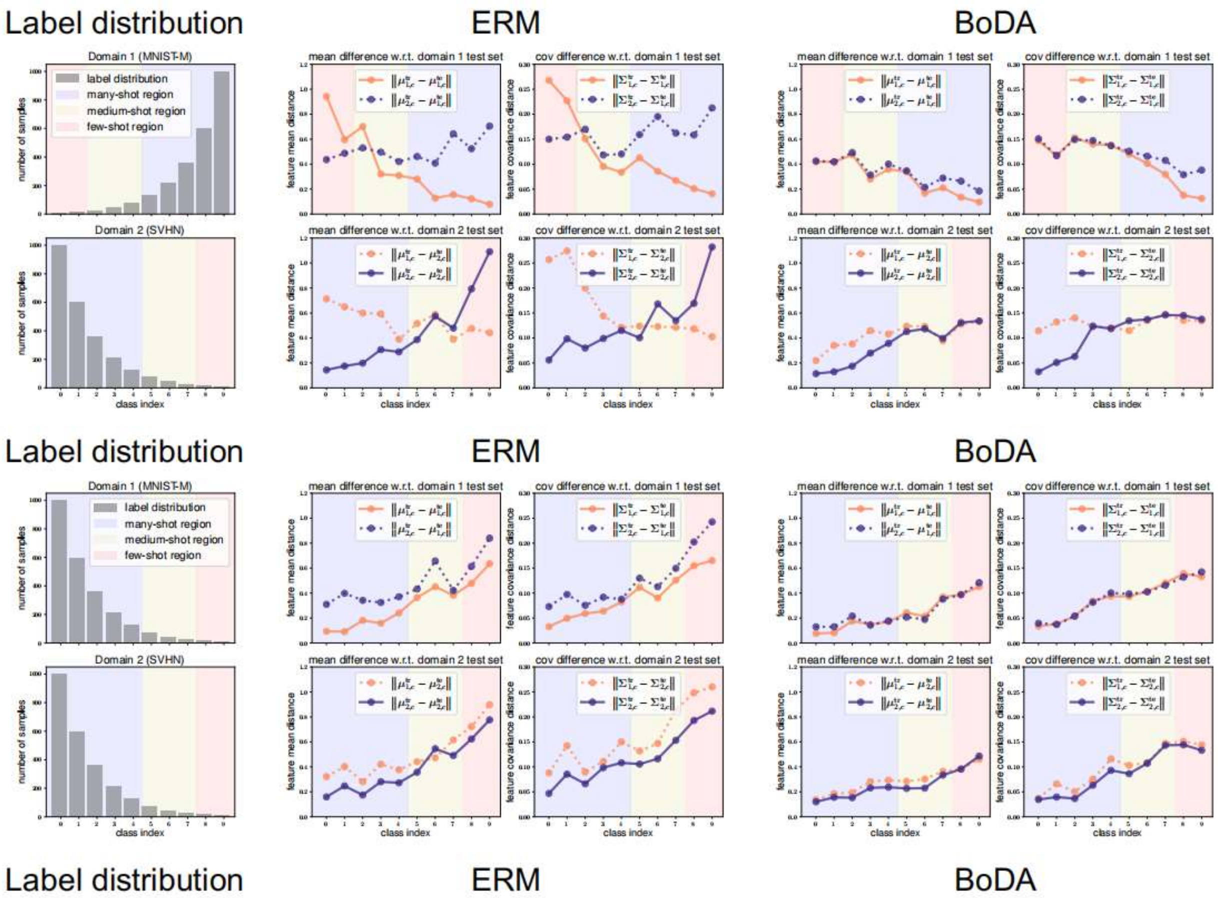
Figure 8: BoDA analysis. **(a)** Label distribution setup. **(b)** Distance of feature mean between train and test data. BoDA enables better learned tail (d, c) with smaller feature discrepancy. **(c)** BoDA learns features that are more aligned across domains even in the presence of divergent labels, and significantly improves upon ERM by 9.5%.

Experiments

Table 9: BoDA strengthens performance on Domain Generalization (DG) benchmarks. Full tables including detailed results for each DG dataset are provided in Appendix G.

Algorithm	VLCS	PACS	OfficeHome	TerraInc	DomainNet	Avg
ERM	77.5 ±0.4	85.5 ±0.2	66.5 ±0.3	46.1 ±1.8	40.9 ±0.1	63.3
Current SOTA [45]	**78.8** ±0.6	86.2 ±0.3	68.7 ±0.3	47.6 ±1.0	41.5 ±0.1	64.5
BoDA$_{r,c}$	78.5 ±0.3	**86.9** ±0.4	**69.3** ±0.1	**50.2** ±0.4	**42.7** ±0.1	**65.5**
BoDA$_{r,c}$ + Current SOTA [45]	79.1 ±0.1	87.9 ±0.5	69.9 ±0.2	50.7 ±0.6	43.5 ±0.3	66.2
BoDA *vs.* ERM	**+1.6**	**+2.4**	**+3.4**	**+4.6**	**+2.6**	**+2.9**

Experiments



Thank you