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

Yuzhe Yang[1] Hao Wang[2] Dina Katabi[1]

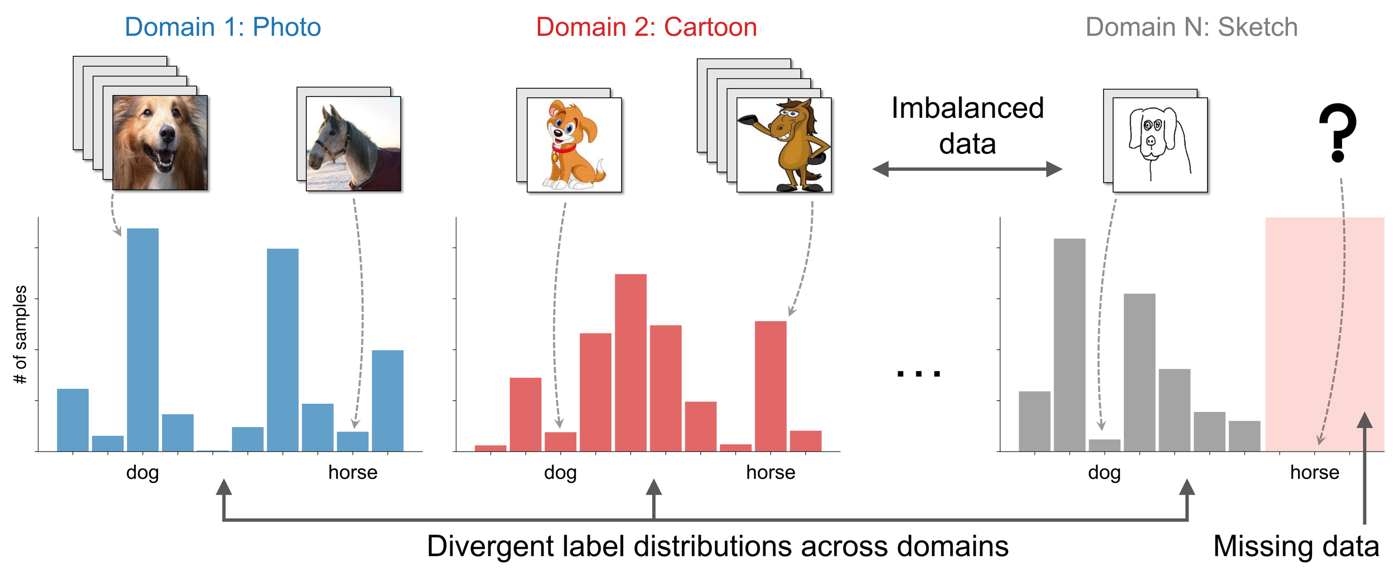
[1]MIT CSAIL [2]Rutgers University

ECCV TEL AVIV 2022

Background & Motivation

1. Existing studies on data imbalance focus on *single domain*
2. Yet, data for one task can originate from distinct *domains*
3. A minority class in one domain can have many samples from other domains, which can help for generalization

Multi-Domain Long-Tailed Recognition (MDLT)

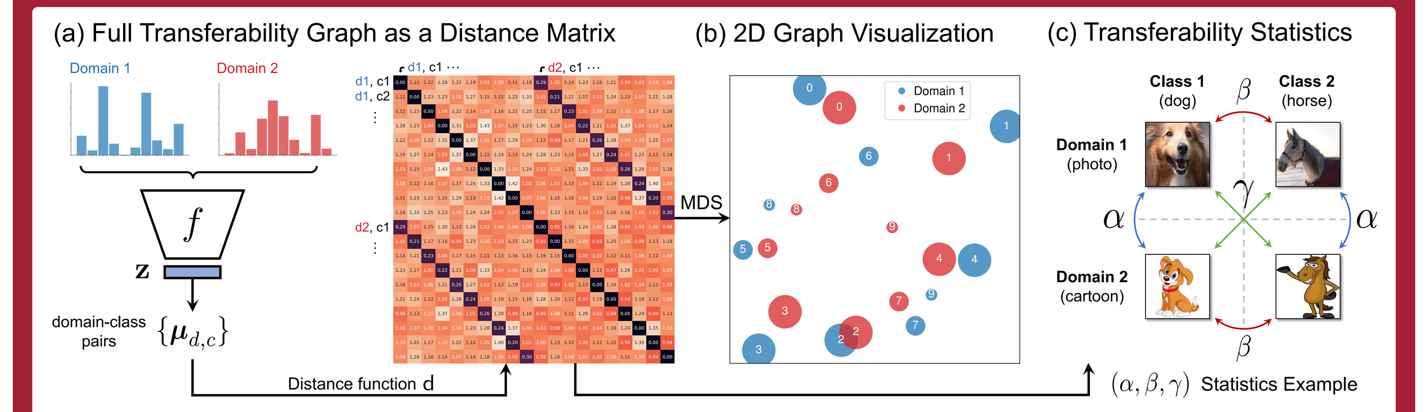


MDLT learns from multi-domain imbalanced data, tackles *label imbalance*, *domain shift*, and *divergent label distributions across domains*, and generalize to ***all*** domain-class pairs

Challenges:

1. *Different label distributions* for each of the domains
2. Multi-domain data inherently involves *domain shift*
3. *Zero-shot* generalization *within* and *across* domains

Domain-Class Transferability Graph



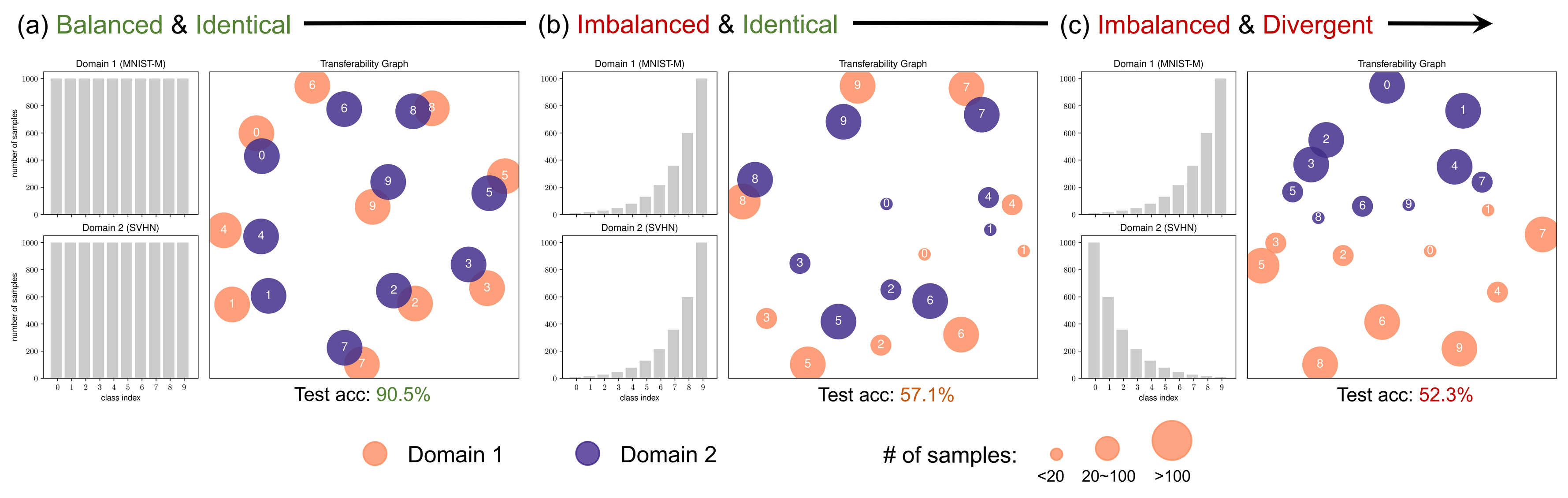
Definition 1: ***Transferability from*** (d, c) ***to*** (d', c')

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

Definition 2: (α, β, γ) ***Transferability Statistics***

$$\textit{Diff domains, same class:} \quad \alpha = \mathbb{E}_c\mathbb{E}_d\mathbb{E}_{d'\neq d}\left[\mathrm{trans}((d,c),(d',c))\right].$$
$$\textit{Same domain, diff classes:} \quad \beta = \mathbb{E}_d\mathbb{E}_c\mathbb{E}_{c'\neq c}\left[\mathrm{trans}((d,c),(d,c'))\right].$$
$$\textit{Diff domains, diff classes:} \quad \gamma = \mathbb{E}_d\mathbb{E}_{d'\neq d}\mathbb{E}_c\mathbb{E}_{c'\neq c}\left[\mathrm{trans}((d,c),(d',c'))\right].$$

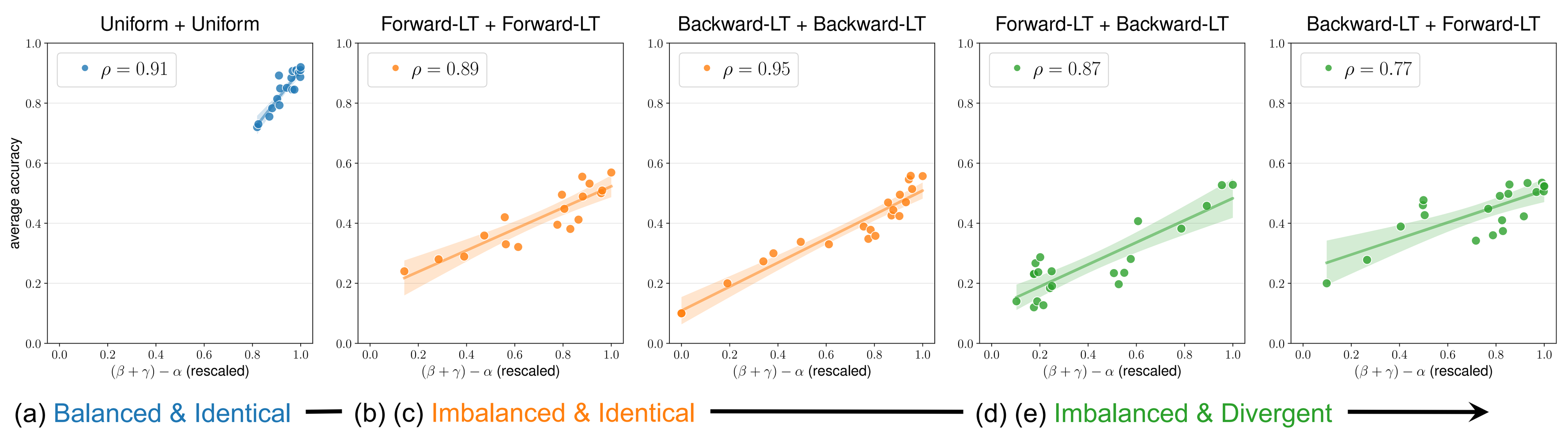
Observation #1: Divergent Label Distributions Hamper Transferable Features



- `Digits-MLT`, a two-domain toy MDLT dataset
- ResNet-18 trained using ERM with different label distributions across domains (gray histograms)
- Different patterns of learned transferability graph:
  1. *Balanced & Identical*: transferable features, high accuracy
  2. *Imbalanced & Identical*: transferable features (majority better than minority classes), moderate accuracy
  3. *Imbalanced & Divergent*: features no longer transferable; clear gap across domains; worst accuracy

Implication: Transferable features needed in MDLT.

Observation #2: Transferability Statistics Characterize Generalization



- Different label configurations for `Digits-MLT`:
  - **Uniform / Forward-LT / Backward-LT**
- 20 ERM models with varying hyperparameters trained for each configuration (each dot a model)
- Plot test accuracy against $(\beta + \gamma - \alpha)$ quantity:
  1. Strong correlation across all ranges / label configurations
  2. Imbalance boosts risk of learning less transferable features

Implication: (α, β, γ) statistics characterize model performance in MDLT.

`BoDA`: A Loss that Bounds the Transferability Statistics

Recall: (α, β, γ) statistics governs the success in MDLT – smaller α and larger β, γ lead to better model performance

A First Approach: Domain-Class Distribution Alignment ($\mathcal{L}_{\mathtt{DA}}$)

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

- ***Pros:*** tackles label *divergence* – numerator $\rightarrow$ *positive* cross-domain pairs (α); denominator $\rightarrow$ *negative* cross-class pairs (β, γ)
- ***Cons:*** does not address label *imbalance* – independent of the number of samples in each (d, c), thus dominated by majority (d, c)

Balanced Domain-Class Distribution Alignment (BoDA).

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

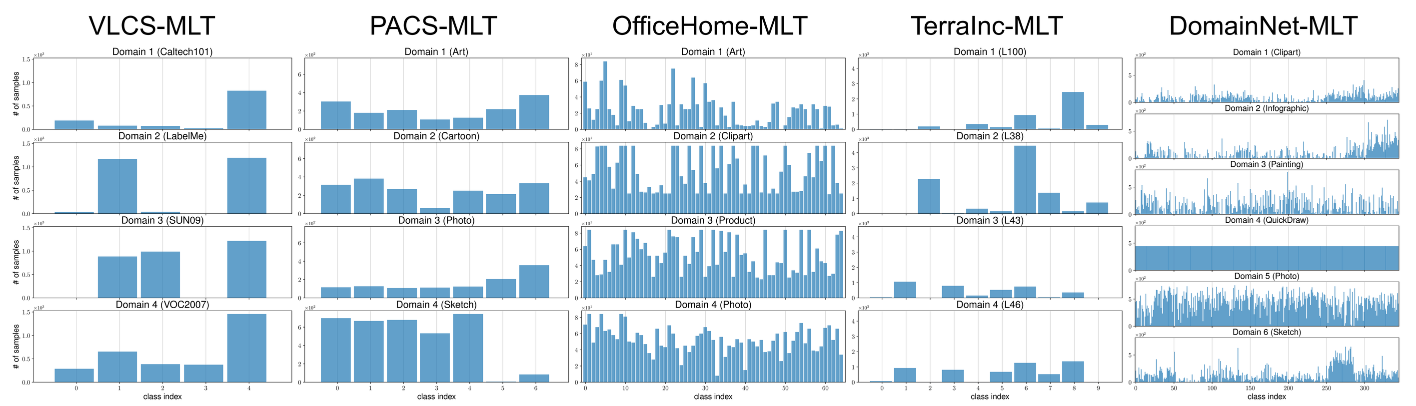
- `BoDA` scales original d by a factor of $1/N_{d_i,c_i}$ – it counters the effect of imbalanced (d, c) by introducing a *balanced* distance $\tilde{\mathsf{d}}$.

Theorem 1 ($\mathcal{L}_{\mathtt{BoDA}}$ 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 (α, β, γ) be the transferability statistics for $\mathcal{S}$. It holds that

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

Implication #1: $\mathcal{L}_{\mathtt{BoDA}}$ upper-bounds (α, β, γ) statistics in a desired form that naturally translates to better performance.
Implication #2: The constant factors correspond to how much each component contributes to the transferability graph.

MDLT Benchmarks + Results



5 MDLT benchmark datasets / ∼20 baseline algorithms

Algorithm	VLCS-MLT	PACS-MLT	OfficeHome-MLT	TerraInc-MLT	DomainNet-MLT	Avg
ERM	76.3 ±0.4	97.1 ±0.1	80.7 ±0.0	75.3 ±0.3	58.6 ±0.2	77.6
Current SOTA	75.9 ±0.5	96.6 ±0.5	81.9 ±0.1	76.4 ±0.5	59.4 ±0.1	78.0
BoDA	**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**

Beyond MDLT: Domain Generalization

- **Domain generalization (DG)**
  - Learn from multiple domains & generalize to *unseen* domains
- **Data imbalance is an *intrinsic problem* in DG**
  1. Learning domains naturally differ in their label distributions
  2. Domains can have (severe) class imbalance within each domain

`BoDA` establishes new SOTA on DG benchmarks

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	**78.8** ±0.6	86.2 ±0.3	68.7 ±0.3	47.6 ±1.0	41.5 ±0.1	64.5
BoDA	78.5 ±0.3	**86.9** ±0.4	**69.3** ±0.1	**50.2** ±0.4	**42.7** ±0.1	**65.5**
BoDA + Current SOTA	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**

Implication: Label imbalance affects out-of-distribution generalization, and is crucial for DG algorithm design.

Conclusion & More Information

New Task	New Techniques	New Benchmarks
MDLT	Domain-Class Trans. Graph BoDA	5 MDLT datasets

Code: `https://github.com/YyzHarry/multi-domain-imbalance`

Project page: `http://mdlt.csail.mit.edu/`