MEnsA: Mix-up Ensemble Average for Unsupervised Multi Target Domain Adaptation on 3D Point Clouds

Ashish Sinha[1] Jonghyun Choi[2]

[1]Simon Fraser University [2]Yonsei University

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Problem Definition

Goal:

Unsupervised MTDA for point cloud via *ensemble average*.

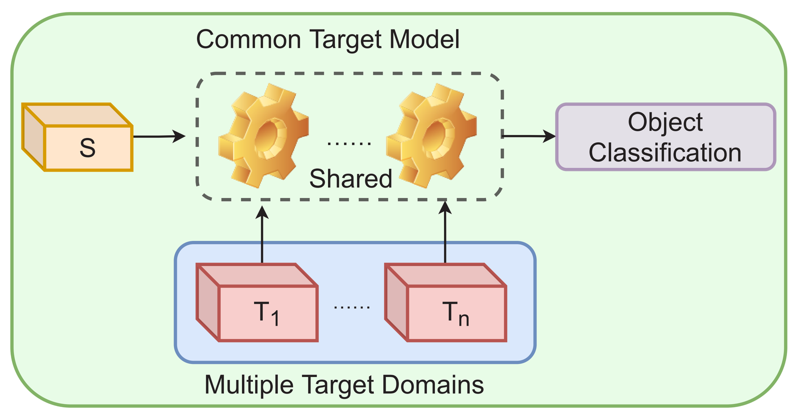
Motivation:

- Prior works focus on STDA for 2D and 3D vision tasks.
- Extending methods of STDA → MTDA is challenging.
- Computational complexity ↑ as # targets ↑.
- A preferred approach is a *single* model for multiple targets.

Contributions

- An ensemble-average mixup approach for MTDA.
- Outperforms previous methods.
- Shows non-generalization ability methods: STDA → MTDA.
- Benchmarks STDA & MTDA methods on point clouds.
- First work on MTDA for 3D data.

MTDA Setup:

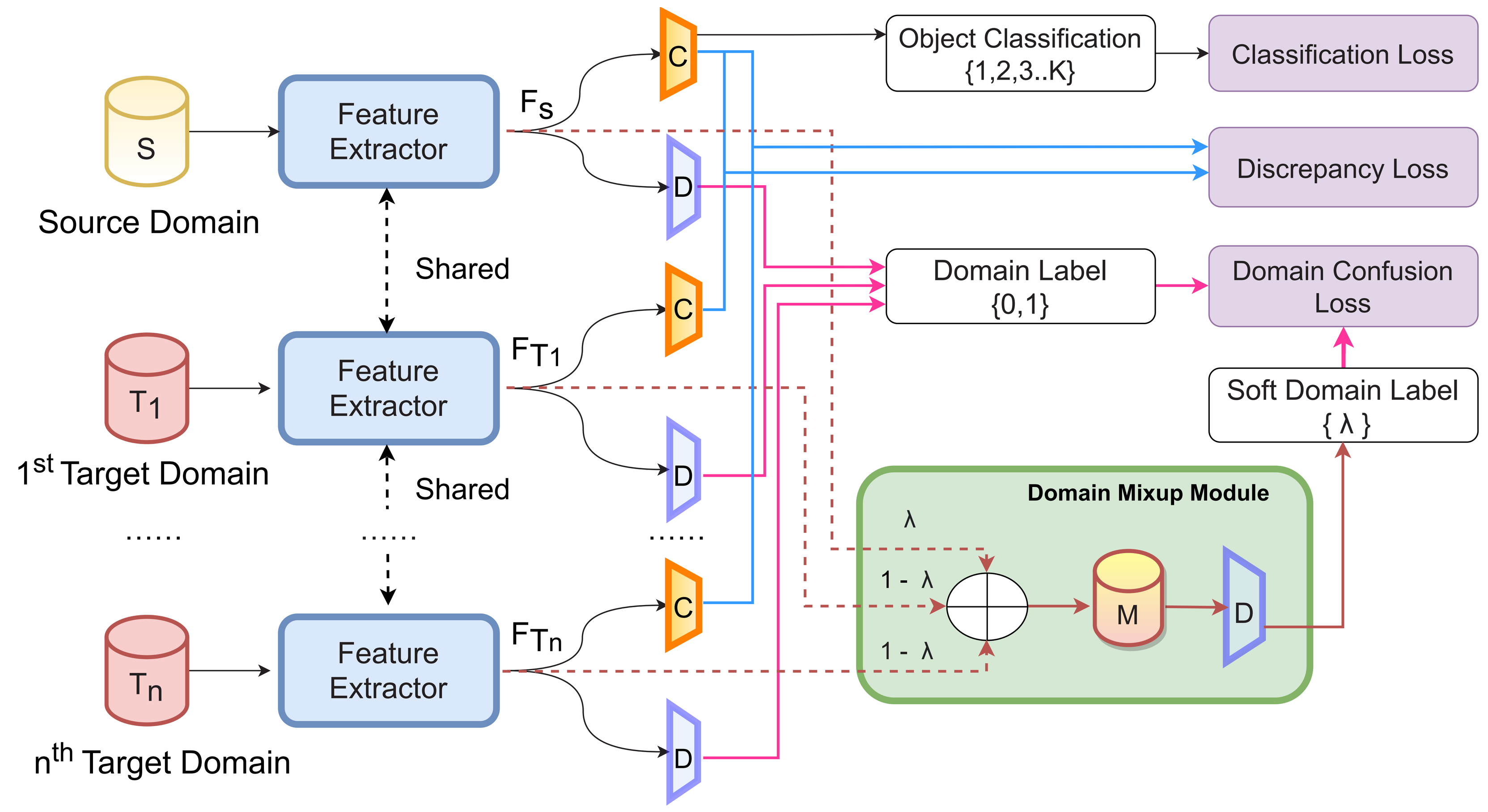


Key Notations

- S/T = Source/ Target Domain
- L_s = Source Domain Label (0)
- L_{T_i} = Target Domain Label (1)
- F_s = Source Embedding
- F_{T_i} = i^{th} Target Embedding
- D = Domain Classifier
- C = Object Classifier
- λ = Mixup Ratio *i.e.* 'Soft' Domain Label

Method

Main Idea: Inspired by mixup, we propose to take an ensemble average of the shared (*i.e.* mixed) latent representations of source and N target domains, modelled as a random variable.

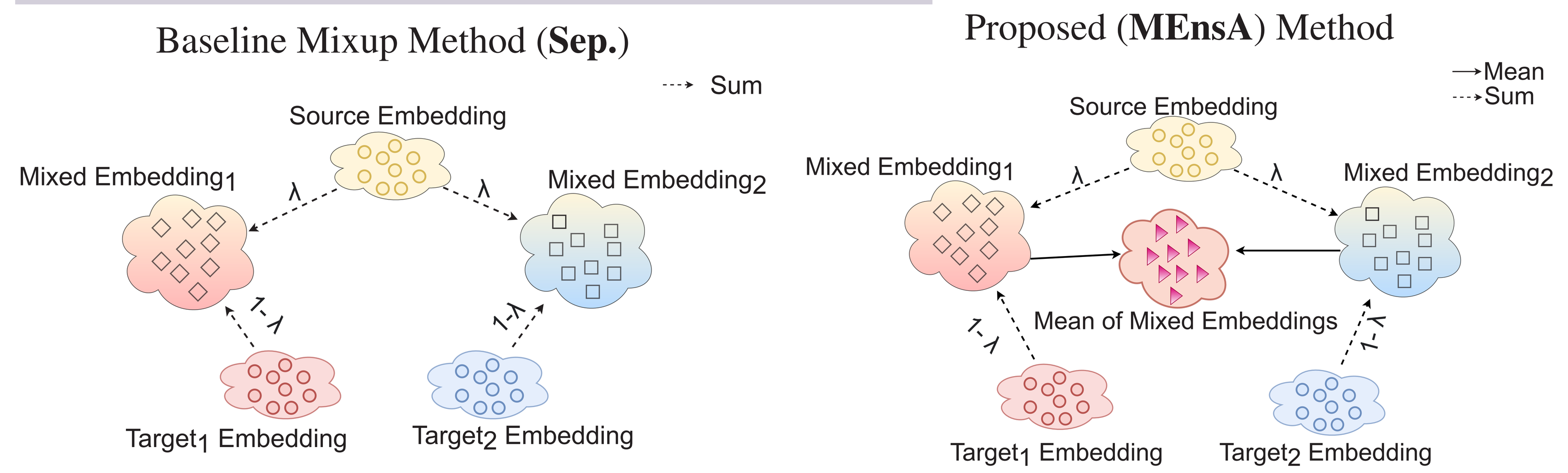


Domain Mixup Module:

Linearly interpolates the source (F_s) and target feature (F_{T_i}) embeddings to obtain F_i^m and the corresponding mixed *soft* domain labels,

$$F_i^m = \lambda F_s + (1-\lambda)F_{T_i}, \quad L_i^m = \lambda L_s + (1-\lambda)L_{T_i}, \quad F_m^M = \frac{1}{n}\sum_{i=1}^{n} F_i^m \quad (1)$$

Comparative Illustration of Mixup Methods:



- Sep. suffers from catastrophic forgetting, as it involves a pair-wise mixup between the source and each target domain.
- MEnsA optimizes the mapping between source and each target domain by reproducing kernel Hilbert space (RKHS) *i.e.* MMD.

Objective Function:

The pipeline is trained end-to-end by minimizing $\mathcal{L}$,

$$\mathcal{L} = \log\left(\sum\left(e^{\gamma(\mathcal{L}_{cls}+\eta\mathcal{L}_{dc}+\zeta\mathcal{L}_{adv})}\right)\right)/\gamma, \quad (5)$$

$$\mathcal{L}_{cls} = \mathcal{L}_{CE}(C(F_s), y_s), \quad \mathcal{L}_{mmd} = \mathcal{L}_{rbf}(C(F_s), F_{T_i}, \sigma),$$

$$\mathcal{L}_{adv} = \lambda_1\mathcal{L}_{mmd} + \lambda_2\mathcal{L}_{dc} + \lambda_3\mathcal{L}_{mixup}, \quad \mathcal{L}_{mixup} = \mathcal{L}_{CE}(D(F_m^M), L_i^m),$$

$$\mathcal{L}_{dc} = \mathcal{L}_{CE}(D(F_s), L_s) + \mathcal{L}_{CE}(D(F_{T_i}, L_{T_i})),$$

Experiments & Results

Quantitative Classification Results (%) on PointDA-10 Dataset

Comparison with prior UDA methods in MTDA setting

Source Domain	ModelNet (M)		ScanNet (S*)		ShapeNet (S)		
Src → Tgt	M → S*	M → S	S* → M	S* → S	S→M	S→S*	Average
No adaptation (Baseline)	35.07	11.75	52.61	29.45	33.65	11.05	28.93
MMD [1]	57.16	22.68	55.40	28.24	36.77	24.88	37.52
DANN [2]	55.03	21.64	54.79	37.37	**42.54**	33.78	40.86
ADDA [3]	29.39	38.46	46.89	20.79	35.33	24.94	32.63
MCD [4]	**57.56**	27.37	54.11	41.71	42.30	22.39	40.94
PointDAN [5]	30.19	44.26	43.17	14.30	26.44	28.92	31.21
AMEAN [6]	55.73	33.53	51.50	30.89	34.73	22.21	38.10
MTDA-ITA [7]	55.23	20.96	56.12	33.71	32.33	25.62	37.33
MT-MTDA [8]	45.43	25.72	28.25	19.51	24.65	**35.27**	29.81
MEnsA (Ours)	45.31	**61.36**	**56.67**	**46.63**	37.02	27.19	**45.70**
↪ w/o mixup	28.48	40.05	33.89	12.14	27.83	24.48	27.81
Supervised in each domain	77.99	67.18	79.83	66.27	63.41	53.02	67.95

Class-wise comparison with prior UDA methods

Method	Bathtub	Bed	Bookshelf	Cabinet	Chair	Lamp	Monitor	Plant	Sofa	Table	Average
No adaptation (Baseline)	40.49	21.95	12.58	6.80	11.11	46.58	51.86	56.00	65.74	46.46	35.96
MMD	55.75	9.75	18.81	0.68	37.54	30.76	46.94	52.00	77.87	75.82	40.59
ADDA	58.71	15.40	23.28	2.68	32.87	50.07	32.95	48.00	61.53	56.6	38.21
DANN	60.42	15.85	24.47	2.72	24.77	12.82	52.03	68.00	65.75	78.42	40.53
MCD	58.72	10.97	27.97	0.68	30.01	12.82	60.33	56.00	82.59	66.06	40.62
AMEAN	58.40	19.05	17.12	7.52	45.17	36.58	54.75	40.00	84.61	72.30	43.55
MTDA-ITA	67.90	11.90	4.11	20.19	21.8	12.19	56.39	45.00	85.38	83.25	40.81
MT-MTDA	59.23	5.88	24.66	4.69	32.08	14.63	66.55	48.00	78.21	72.66	40.66
MEnsA (Ours)	67.11	6.58	6.77	44.89	74.09	46.05	87.92	64.55	50.00	74.47	52.24
Supervised in each domain	91.10	69.51	61.05	89.23	99.67	80.76	91.57	51.37	94.08	81.97	81.03

Comparison with Variants of our Mixup Formulation

Source Domain	ModelNet (M)		ScanNet (S*)		ShapeNet (S)		
Src → Tgt	M → S*	M →S	S* → M	S*→ S	S→M	S→S*	Average
MEnsA (Ours)	45.31	**61.36**	**56.67**	**46.63**	**37.02**	27.19	**45.70**
Mixup Sep	41.32	47.98	56.18	42.19	28.85	36.69	42.20
Factor-Mixup	41.31	41.49	50.77	38.82	30.77	36.81	40.00
Concat-Mixup	49.20	29.57	50.47	37.5	33.05	25.64	37.57
Inter-Mixup	**50.95**	28.65	51.71	34.38	32.21	**40.80**	39.78
Best of all methods	50.95	61.36	56.67	46.63	37.02	40.80	48.91

Ablation on Contribution of each Module

Source Domain	ModelNet (M)		ScanNet (S*)		ShapeNet (S)		
Loss Terms (Eq. 5)	M → S*	M →S	S* → M	S* → S	S→M	S→S*	Average
$\mathcal{L}_{dc}$	34.42	45.08	32.81	13.32	23.55	**38.13**	31.22
$\mathcal{L}_{mmd}$	43.37	36.05	51.87	29.20	30.67	25.75	36.15
$\mathcal{L}_{mix}$	32.67	43.51	**57.88**	33.17	30.52	31.59	38.22
$\mathcal{L}_{dc} + \mathcal{L}_{mmd}$	41.05	41.78	42.67	19.83	29.08	33.62	34.67
$\mathcal{L}_{dc} + \mathcal{L}_{mix}$	35.07	45.19	35.29	16.34	22.59	26.79	30.21
$\mathcal{L}_{mmd} + \mathcal{L}_{mix}$	43.47	53.17	55.95	30.04	28.60	30.40	40.27
$\mathcal{L}_{dc} + \mathcal{L}_{mix} + \mathcal{L}_{mmd}$	**45.31**	**61.36**	56.67	**46.63**	**37.02**	27.19	**45.70**

Variants of Our Mixup Formulation

Factor-Mixup:

We consider the effect of *scaling* factor while ensembling in Eq. 1.

$$F_m^{factor} = \lambda F_s + \sum_{i=1}^{n}\frac{1-\lambda}{n}F_{T_i}$$

Concat-Mixup:

We consider how Eq. 1 is affected when using *concatenation* instead of summation.

$$F_m^{concat} = [\lambda F_s, \frac{1-\lambda}{n}F_{T_1}, ..., \frac{1-\lambda}{n}F_{T_n}],$$

$$L_m^{concat} = [\lambda, 2\frac{1-\lambda}{n}, .., N\frac{1-\lambda}{n}]$$

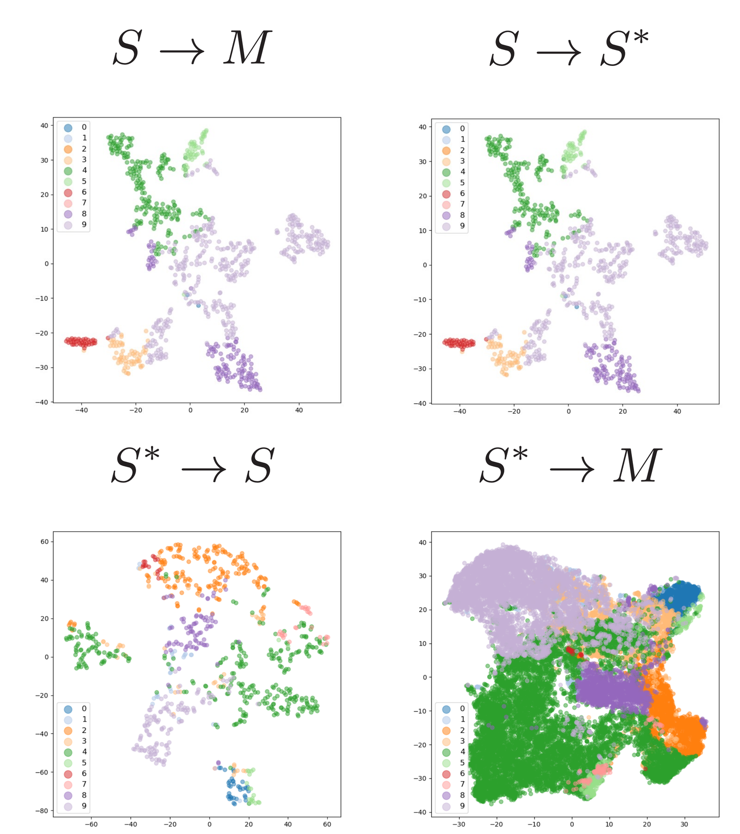
Inter-Mixup:

Here, we consider extending the linear interpolation as per Eq. 1 by incorporating the target domains as well.

$$F_m^T = \lambda F_{T_1} + (1-\lambda)F_{T_2},$$

$$L_m^T = \lambda L_{T_1} + (1-\lambda)L_{T_2}.$$

T-SNE Embedding Visualizations



References:

[1] MMD [Long *et al.*, ICCV '13]
[2] DANN [Ganin *et al.*, PMLR '15]
[3] ADDA [Zheng *et al.*, CVPR '17]
[4] MCD [Saito *et al.*, CVPR '18]
[5] PointDAN [Qin *et al.*, NeurIPS'19]
[6] AMEAN [Chen *et al.*, CVPR '19]
[7] MTDA-ITA [Gholami *et al.*, TIP '20]
[8] MT-MTDA [Nguyen *et al.*, WACV '21]

Code: