Deep Joint Distribution Optimal Transport for Universal Domain Adaptation on Time Series

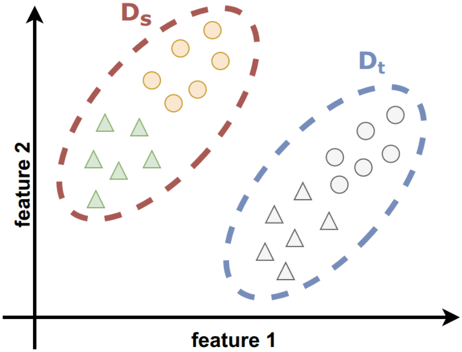
IJCNN 2025

Romain Mussard, Fannia Pacheco, Maxime Berar, Gilles Gasso & Paul Honeine

Univ Rouen Normandie, INSA Rouen Normandie, Normandie Univ, LITIS UR 4108

The Challenge of Universal Domain Adaptation

(Unsupervised) Domain Adaptation

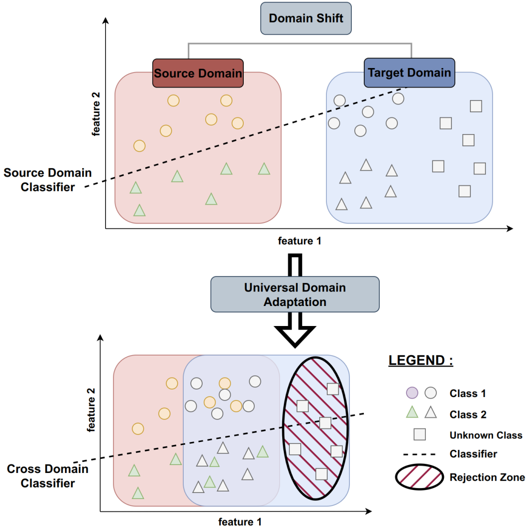


Covariate Shift

Datasets & Distributions

- Source Domain : $\mathcal{D}_s = \{x_i^s, y_i^s\}_{i=0}^{n_s} \sim P_s(x, y)$
- Target Domain : $\mathcal{D}_t = \{x_i^t\}_{i=0}^{n_t} \sim P_t(x, y)$
- $P_s(x) \neq P_t(x)$

Universal Domain Adaptation



Objectives :

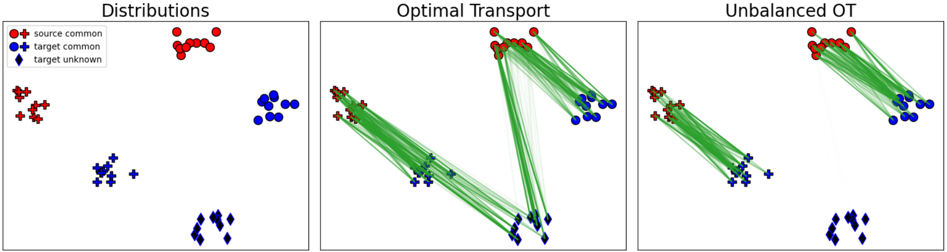
- **Alignment:** Align source and target common samples
- **OOD Discovery:** Detect target private samples (OOD samples)

No prior assumption is made regarding target set distribution !

UniDA State-of-the-Art

Method	Alignment	OOD Discovery	Threshold Based
UAN [1]	Adversarial Discriminator	Non Adversarial Discriminator	✓
DANCE [2]	Neighborhood Clustering	Entropy Separation	✓
OVANet [3]	Entropy	One-vs-All Rejection	✗
RAINCOAT [4]	Wasserstein	Statistic Test	✓
UniOT [5]	Neighborhood Clustering	Optimal Transport	✓
PPOT [6]	Prototypical Wasserstein	Softmax Thresholding	✓

Discrete Optimal Transport



When $\mu_s = \sum_{i=1}^{n_s} \mathbf{a}_i \delta_{\mathbf{x}_i^s}$ and $\mu_t = \sum_{j=1}^{n_t} \mathbf{b}_j \delta_{\mathbf{x}_j^t}$

$$\gamma_0 = \underset{\gamma \in \Pi(\mathbf{a},\mathbf{b})}{\operatorname{argmin}} \left\{ \langle \gamma, \mathbf{C} \rangle = \sum_{i,j} \gamma_{i,j} c_{i,j} \right\}$$

s.t $\mathbf{C}$ is a cost matrix with $c_{i,j} = c(\mathbf{x}_i^s, \mathbf{x}_j^t)$ and the marginals constraints are :

$$\Pi(\mathbf{a}, \mathbf{b}) = \left\{ \gamma \in \mathbb{R}_+^{n_s \times n_t} \mid \gamma \mathbf{1}_{n_t} = \mathbf{a}, \gamma^\top \mathbf{1}_{n_s} = \mathbf{b} \right\}$$

***p*-Wasserstein distance** : $W^p(\mathbf{a}, \mathbf{b}) = \langle \gamma_0, \mathbf{C} \rangle^{1/p}$.

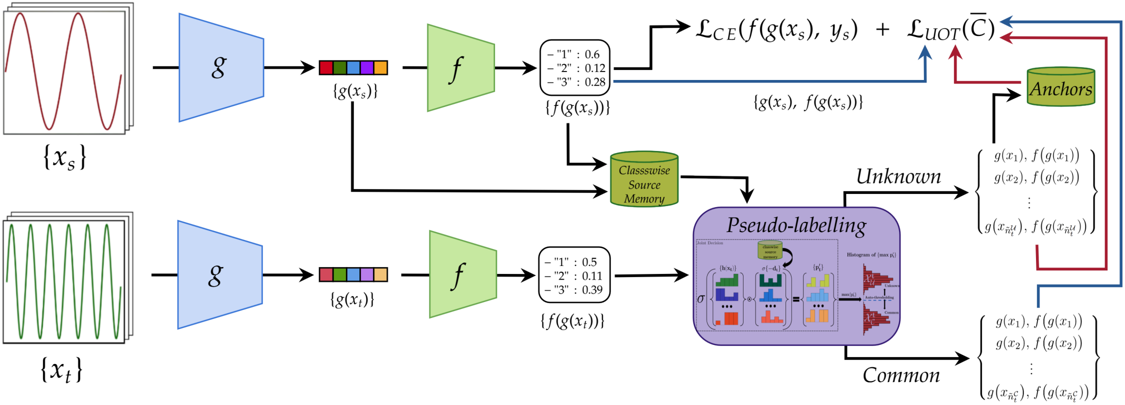
Relaxed OT :

- **Partial OT:** Match only a portion of the total mass: $(1_{n_s}^\top \gamma 1_{n_t} = m \leq 1)$
- **Unbalanced OT:** Allow for mass creation and destruction. $(\gamma > 0)$

Source : [7]

UniJDOT : Adaptive Thresholding and Joint Decision Strategy

Universal Joint Distribution Optimal Transport (UniJDOT)



UniJDOT Architecture

Unknown Target Detection

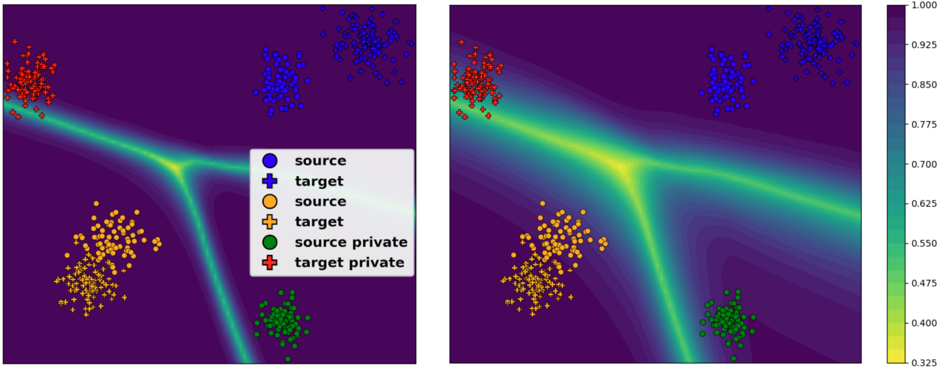
Joint Decision : Thresholding over a feature space conditioning of the classifier's logits

$$\max \sigma\left(h(x^t)\sigma(-d_t)\right) < \tau,$$

With d the minimal distance in the feature space between target point x^t and closest source point x^s of each source class :

$$d_t = \left(\min_{x^s \in \mathcal{X}_s^1} d(x^t, x^s), \cdots, \min_{x^s \in \mathcal{X}_s^K} d(x^t, x^s)\right).$$

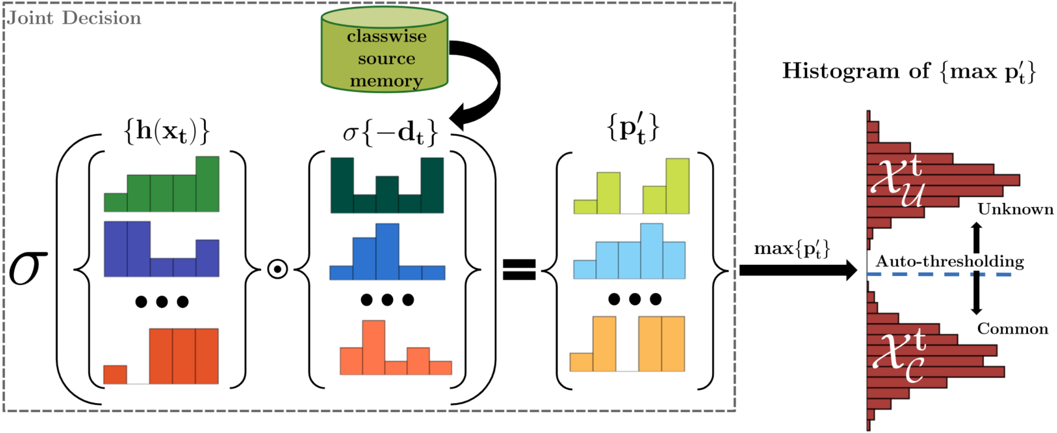
Joint Decision Space



(a) Softmax Decision (b) Joint Decision

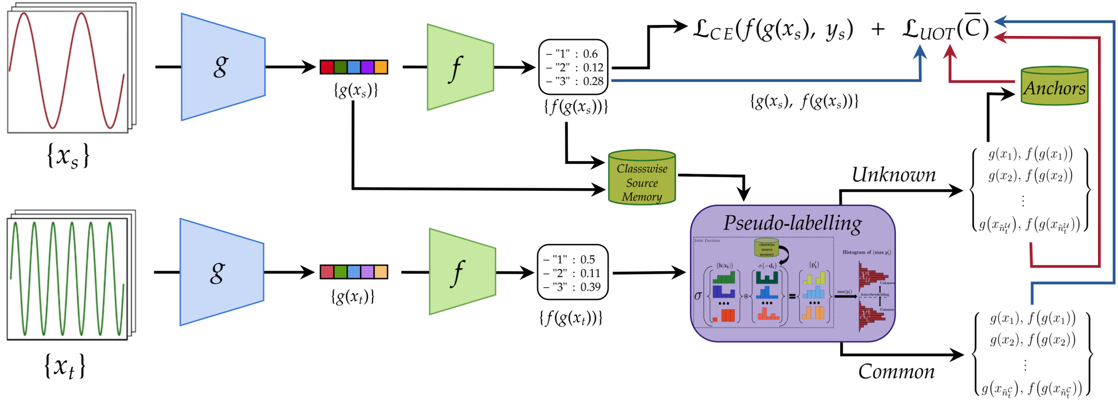
Decision space of a 2D toy dataset

Adaptive Pseudo-labelling



Pseudo-labelling

Universal Joint Distribution Optimal Transport (UniJDOT)



UniJDOT Architecture

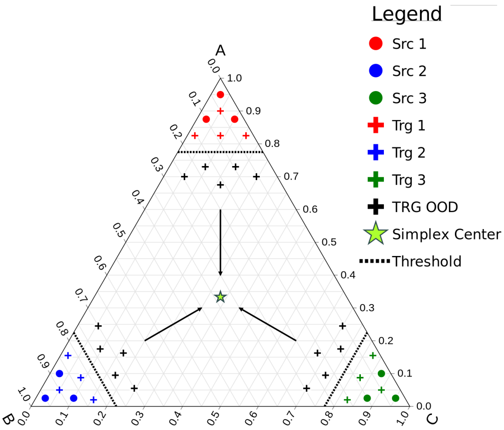
Cost and Attractors

Any $x_j^t \in \mathcal{X}_{\mathcal{C}}^t$ should be aligned with the source samples:

$$\mathbf{C}_{ij}^{\mathcal{C}} = \lambda_1 \|g(x_i^s) - g(x_j^t)\|_2^2 + \lambda_2 \|y_i^s - f(g(x_j^t))\|_2^2,$$

Any $x_j^t \in \mathcal{X}_{\mathcal{O}}^t$, should be aligned with OOD attractors:

$$\mathbf{C}_{\cdot j}^{\mathcal{O}} = \lambda_1 \|a - g(x_j^t)\|_2^2 + \lambda_2 \|r - f(g(x_j^t))\|_2^2,$$



Simplex Attractor

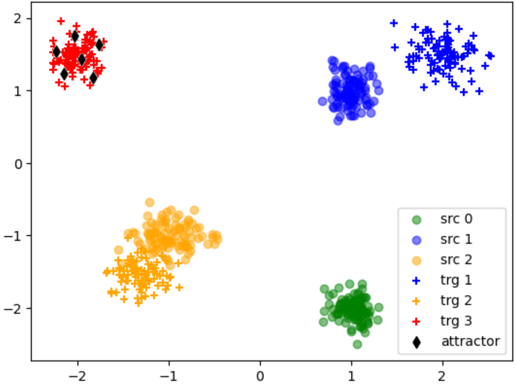
Cost and Attractors

Any $x_j^t \in \mathcal{X}_{\mathcal{C}}^t$ should be aligned with the source samples:

$$\mathbf{C}_{ij}^{\mathcal{C}} = \lambda_1 \|g(x_i^s) - g(x_j^t)\|_2^2 + \lambda_2 \|y_i^s - f(g(x_j^t))\|_2^2,$$

Any $x_j^t \in \mathcal{X}_{\mathcal{O}}^t$, should be aligned with OOD attractors:

$$\mathbf{C}_{\cdot j}^{\mathcal{O}} = \lambda_1 \|a - g(x_j^t)\|_2^2 + \lambda_2 \|r - f(g(x_j^t))\|_2^2,$$



Feature Attractor

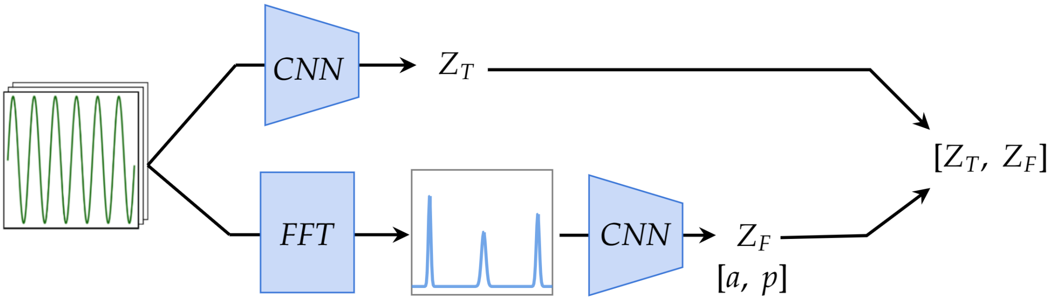
Rewriting the cost matrix C :

$$\bar{\mathbf{C}} = \begin{array}{c} \\ \mathcal{X}^s \\ \mathcal{A} \end{array} \begin{array}{c} \begin{array}{cc} \mathcal{X}_t^{\mathcal{C}} & \mathcal{X}_t^{\mathcal{O}} \end{array} \\ \begin{bmatrix} \mathbf{C}^{\mathcal{C}} & \xi^{S\mathcal{O}} \\ \xi^{A\mathcal{C}} & \mathbf{C}^{\mathcal{O}} \end{bmatrix} \end{array}, \tag{1}$$

Aligning using UOT :

$$\mathbf{UOT}(\mathbf{a}, \mathbf{b}, \bar{\mathbf{C}}) = \min_{\gamma \geq 0} \langle \bar{\mathbf{C}}, \gamma \rangle + \tau_1 KL(\gamma 1_{n_t}, \mathbf{a}) + \tau_2 KL(\gamma^\top 1_{n_s}, \mathbf{b}), \tag{2}$$

Frequency Feature



Time-Frequency Feature Extractor

Experimental Results

UniDA Results

Scenario	UAN	OVANet	PPOT*	DANCE	UniOT*	UniJDOT*
12 → 16	**57 ± 06**	19 ± 19	53 ± 10	34 ± 25	30 ± 15	50 ± 13
13 → 3	72 ± 04	38 ± 31	53 ± 37	**85 ± 10**	69 ± 05	69 ± 11
15 → 21	77 ± 19	30 ± 32	52 ± 39	**91 ± 06**	77 ± 02	75 ± 06
17 → 29	69 ± 07	17 ± 28	**77 ± 11**	71 ± 25	71 ± 03	73 ± 05
1 → 14	**80 ± 04**	06 ± 10	48 ± 25	07 ± 12	64 ± 21	44 ± 33
22 → 4	74 ± 06	48 ± 25	61 ± 34	**82 ± 02**	67 ± 06	71 ± 08
24 → 8	41 ± 12	09 ± 17	**59 ± 08**	58 ± 11	55 ± 11	47 ± 20
30 → 20	37 ± 11	20 ± 18	49 ± 17	19 ± 27	34 ± 14	**50 ± 07**
6 → 23	24 ± 10	29 ± 29	**76 ± 07**	08 ± 26	53 ± 14	70 ± 05
9 → 18	**69 ± 09**	42 ± 28	57 ± 08	53 ± 13	49 ± 07	66 ± 08
mean	60	26	59	51	57	**62**

H-score (%) for HAR dataset

Scenario	UAN	OVANet	PPOT*	DANCE	UniOT*	UniJDOT*
0 → 2	**47 ± 17**	20 ± 14	01 ± 01	20 ± 26	42 ± 17	40 ± 19
0 → 6	43 ± 10	41 ± 12	07 ± 04	46 ± 28	**56 ± 10**	48 ± 11
1 → 6	44 ± 15	21 ± 21	05 ± 02	64 ± 15	51 ± 20	**66 ± 11**
2 → 7	**41 ± 08**	22 ± 12	10 ± 03	20 ± 21	10 ± 09	28 ± 06
3 → 8	58 ± 20	52 ± 26	03 ± 03	69 ± 21	55 ± 12	**71 ± 12**
4 → 5	48 ± 16	17 ± 18	02 ± 02	02 ± 02	40 ± 17	**57 ± 18**
5 → 0	16 ± 09	08 ± 05	02 ± 01	00 ± 01	**21 ± 08**	17 ± 14
6 → 1	77 ± 14	31 ± 25	01 ± 02	62 ± 42	72 ± 16	**79 ± 13**
7 → 4	45 ± 10	16 ± 15	04 ± 03	06 ± 04	62 ± 10	**64 ± 19**
8 → 3	42 ± 31	32 ± 22	01 ± 01	**69 ± 33**	32 ± 20	56 ± 18
mean	46	26	03	36	44	**53**

H-score (%) for HHAR dataset

Scenario	UAN	OVANet	PPOT*	DANCE	UniOT*	UniJDOT*
0 → 11	**37 ± 13**	28 ± 10	05 ± 09	24 ± 18	18 ± 14	35 ± 13
12 → 5	51 ± 10	32 ± 19	27 ± 09	54 ± 13	60 ± 08	**65 ± 05**
13 → 17	32 ± 06	31 ± 12	26 ± 07	50 ± 15	39 ± 15	**54 ± 11**
16 → 1	45 ± 05	40 ± 14	31 ± 07	25 ± 12	37 ± 03	**50 ± 05**
18 → 12	31 ± 04	28 ± 14	18 ± 10	20 ± 07	27 ± 04	**33 ± 04**
3 → 19	37 ± 03	**45 ± 16**	23 ± 06	39 ± 18	38 ± 05	42 ± 10
5 → 15	36 ± 10	53 ± 12	16 ± 06	42 ± 27	**66 ± 03**	61 ± 04
6 → 2	**55 ± 02**	36 ± 10	30 ± 11	25 ± 06	33 ± 04	42 ± 04
7 → 18	53 ± 02	47 ± 17	36 ± 07	31 ± 11	55 ± 05	**56 ± 02**
9 → 14	43 ± 04	53 ± 21	28 ± 06	62 ± 16	64 ± 06	**70 ± 05**
mean	42	39	24	37	44	**51**

*Models trained with CNN+FNO

H-score (%) for EDF dataset

$$\textbf{H-score} = \frac{2A_cA_u}{A_c + A_u}$$

A_c is the accuracy of known classes.

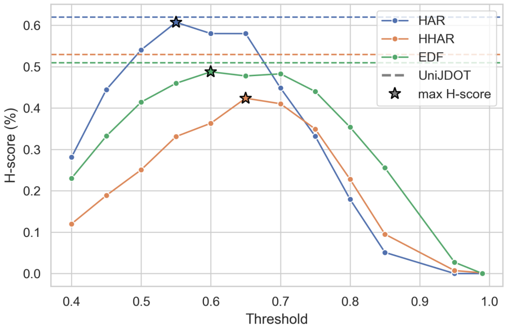
A_u is the accuracy of target unknown classes.

Ablation Study

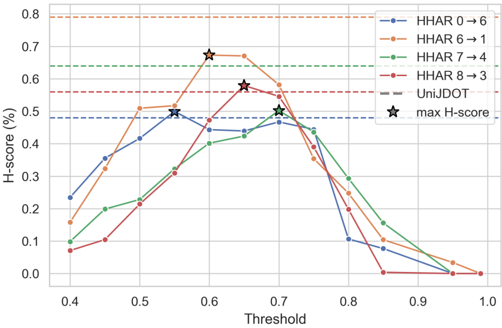
Ablation			Datasets		
Auto-Thresh.	Joint Decision	FNO	HAR	HHAR	EDF
✓	✓	✓	**62**	<u>53</u>	**51**
✓	✓	✗	52	**59**	47
✓	✗	✓	53	40	45
✓	✗	✗	39	43	42
✗	✓	✓	<u>61</u>	41	<u>49</u>
✗	✓	✗	19	39	43
✗	✗	✓	2	4	14
✗	✗	✗	1	5	12

Ablation study (H-scores)

Threshold Sensitivity



(a) Interdataset threshold sensitivity



(b) Intradataset threshold sensitivity

Threshold Sensitivity

Conclusion

Wrap-Up:

- UniJDOT demonstrate SOTA results over Time Series Classification
- UniJDOT introduce a combination of joint decision and adaptive thresholding

Future Work:

- **Focus on Time Series :** How to learn a good representation of such data?

Thank You!

romain.mussard@univ-rouen.fr

Questions?

References i

[1] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Universal domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[2] K. Saito, D. Kim, S. Sclaroff, and K. Saenko, "Universal domain adaptation through self-supervision," 2020.

[3] K. Saito and K. Saenko, "Ovanet: One-vs-all network for universal domain adaptation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8980–8989, 2021.

[4] H. He, O. Queen, T. Koker, C. Cuevas, T. Tsiligkaridis, and M. Zitnik, "Domain adaptation for time series under feature and label shifts," in *International Conference on Machine Learning*, pp. 12746–12774, PMLR, 2023.

[5] W. Chang, Y. Shi, H. Tuan, and J. Wang, "Unified optimal transport framework for universal domain adaptation," in *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 29512–29524, Curran Associates, Inc., 2022.

[6] Y. Yang, X. Gu, and J. Sun, "Prototypical partial optimal transport for universal domain adaptation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 10852–10860, Jun. 2023.

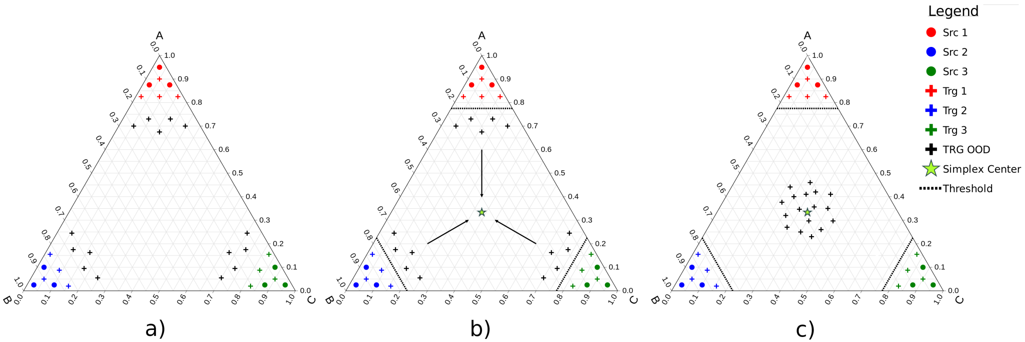
[7] N. C. Rémi Flamary, "Optimal transport for machine learning," 2019. `https://remi.flamary.com/cours/otml/OTML_ISBI_2019_OTintro.pdf` **[Accessed: May 30, 2024].**

Auto-Thresholding Selection

Comparison of auto-thresholding methods
(H-scores)

Datasets	Auto-thresholding Methods			
	Yen	Otsu	Triangle	Li
HAR	**62**	<u>55</u>	41	50
HHAR	**53**	<u>40</u>	34	38
EDF	**51**	<u>48</u>	30	47

Illustration of UniJDOT



a) Simplex of a pretrained model over source domain for a 3-source-class problem. b) the objective of UniJDOT is to push unknown samples in the simplex center. c) Expected simplex space after UniJDOT training.

Fourier Neural Operators (FNO)

Dataset	UniOT	UniOT-FNO	UniJDOT	UniJDOT-FNO
HAR	44	51	50	**54**
HHAR	43	47	47	**50**

H-score w/o FNO : UniOT-FNO and UniJDOT-FNO are using FNO as feature extractor

Fourier Neural Operator (FNO) :

1) Smooth: $x_i = smooth(x_i)$
2) DFT: $v_i = DFT(x_i)$
3) Convolution: $\tilde{v}_i = B * v_i$
4) transform: $a_i, p_i \leftarrow \tilde{v}_i$
5) Extract: $e_{F,i} = [a_i, p_i]$