Towards Uncertainty-aware Unsupervised Domain Adaptation for Videos and Time-Series with Causal Optimal Transport

Khushboo Mishra, Varun Trivedi, Tanima Dutta

CVPR
JUNE 3-7, 2026
DENVER
COLORADO

Department of Computer Science and Engineering
Indian Institute of Technology (BHU) Varanasi, India

Contents

Introduction

What is Domain Adaptation ?

- Train model on source domain (labeled) and deploy on target domain (unlabeled).
- Goal: Generalize across domains.

Why Domain Adaptation is Needed?

Domain Shift: $P_s(x, y) \neq P_t(x, y)$

- Leads to poor generalization and performance drop in real-world data

Challenges in Deep Neural Networks

- Learn domain-specific features
- Sensitive to noise, style variations, and distribution changes $\Rightarrow$ poor transferability across domains

Why Optimal Transport?

- Optimal Transport aligns distributions geometrically while preserving structure and enabling nonlinear mapping.

Example of Domain Adaptation

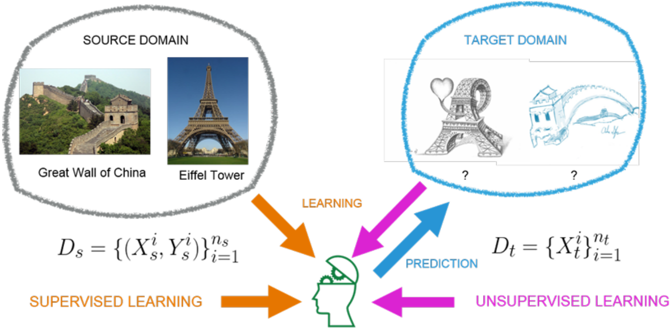

Figure: Leveraging labeled source domain, to learn a model for the target domain.

Motivation of the Work

Motivation

- Domain adaptation methods are typically designed under **simplified assumptions**, such as linear shifts and shared label spaces.
- In real-world scenarios, data exhibits **nonlinear distribution shifts, noise, and high variability**, which degrade model performance.
- The presence of **unknown target classes** leads to negative transfer and incorrect alignment.
- In time-series data, **non-stationarity and prediction uncertainty** further complicate reliable adaptation.
- These challenges motivate the need for a **robust Optimal Transport-guided framework** that ensures reliable and transferable domain adaptation.

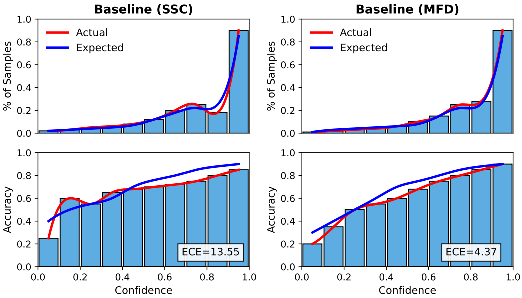

Figure: Miscalibrated predictions under temporal domain shifts

- Baseline shows high calibration error (ECE)
- Proposed method reduces ECE $\rightarrow$ better alignment

Observation

- Temporal domain shifts lead to **miscalibrated predictions**
- Results in **overconfident and unreliable decisions**

Proposed Approach

To solve the problem:

We propose Causal-OT, a novel UDA framework that integrates **causal graph regularization** into **optimal transport** to preserve domain-invariant causal structures during alignment. Additionally, we introduce a **causality-aware pseudo-labeling** strategy that selects reliable target samples based on confidence and structural consistency, improving robustness against noisy pseudo-labels.

Why Causal-OT?

Method	Dist.	Causal	Uncertainty
TransPL	✓	×	×
CauDiTS	×	✓	×
RAINCOAT	✓	×	×
Causal-OT	✓	✓	✓

Only method addressing all key challenges

Our Solution

Contributions

- **Unified Causal-OT Framework:** Propose a novel UDA framework that jointly handles **distribution alignment, causal structure, and prediction uncertainty** for robust domain adaptation.
- **Causal Structure-Preserving OT:** Integrate **Granger causal graphs** into OT to align feature distributions while preserving inter-channel causal dependencies.
- **Uncertainty-Aware Pseudo-Labeling:** Introduce a **causality-aware filtering strategy** based on entropy and structural consistency to reduce pseudo-label noise.
- **Strong Empirical Performance:** Achieve consistent improvements on time-series and video benchmarks, with up to **4.5% accuracy** and **3.8% F1-score gains**.

Overall Architecture

Proposed Architecture

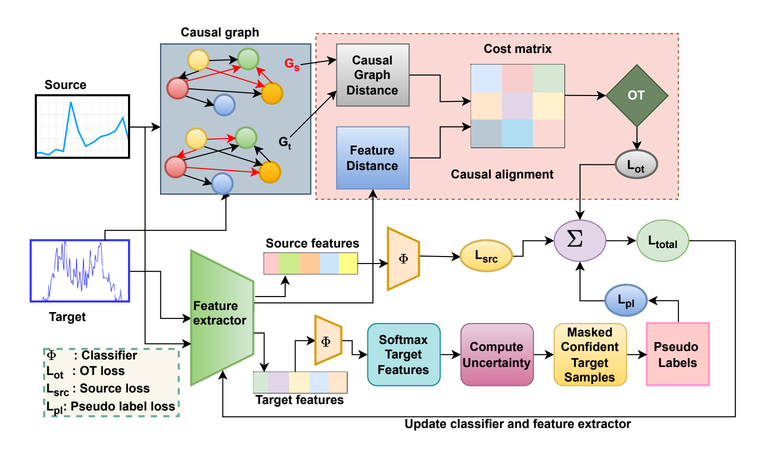

Figure: Overview of the proposed Causal-OT framework.

Causal Graph Construction

Granger Causal Graphs

- Extract causal relations from time-series using Granger causality
- Construct graph:

$$G = (V, E, W)$$

- Captures inter-variable temporal dependencies

$X \in \mathbb{R}^{N \times T \times D}$: Input data
V : Variables (nodes)

W : Adjacency matrix
G_s, G_t : Source/Target graphs

Causal Feature Embedding

Graph Embedding

$$L = D - W$$

$$\Phi(G) \in \mathbb{R}^{d \times k}$$

- Extract causal descriptors:

$$\phi_i = \Phi(G)[i]$$

L : Graph Laplacian
$\Phi(G)$: Embedding

ϕ_i : Node descriptor
k : Embedding dimension

Causal OT Cost

Cost Function

$$C_{ij} = \|z_i^s - z_j^t\|^2 + \lambda\|\phi_i^s - \phi_j^t\|^2$$

- Combines feature + causal similarity

C_{ij} : Transport cost
z_i^s, z_j^t : Features

ϕ_i, ϕ_j : Causal descriptors
λ : Weight

Optimal Transport Alignment

OT Objective

$$\gamma^* = \arg\min_{\gamma} \langle \gamma, C \rangle + \varepsilon H(\gamma)$$

- Align distributions with causal awareness

γ : Transport plan
C : Cost matrix

$H(\gamma)$: Entropy
ε : Regularization

Uncertainty Estimation

Entropy-Based Uncertainty

$$U_t^j = -\sum_k \hat{y}_{t,j}^{(k)} \log \hat{y}_{t,j}^{(k)}$$

- High entropy $\rightarrow$ uncertain prediction

$\hat{y}_t$: Prediction
K : Classes

U_t : Uncertainty
j : Sample index

Uncertainty-Aware Pseudo-Labeling

Filtering Strategy

$$I = \{j \mid U_t^j < \rho\}$$

- Use only confident samples

$$\mathcal{L}_{PL} = \sum_{j \in I} \mathrm{CE}(h_\phi(z_t^j), \hat{y}_t^j)$$

I : Selected samples
ρ : Threshold

$\mathcal{L}_{PL}$: Loss
h_ϕ : Classifier

Overall Objective

Training Loss

$$\mathcal{L} = \mathcal{L}_{src} + \alpha\mathcal{L}_{OT} + \beta\mathcal{L}_{PL}$$

$\mathcal{L}_{src}$: Source loss
$\mathcal{L}_{OT}$: Alignment

$\mathcal{L}_{PL}$: Pseudo-label
α, β : Weights

Comparison with Existing Methods on UCF101 $\leftrightarrow$ HMDB51

Quantitative Comparison on UCF101 $\leftrightarrow$ HMDB51

Method	**U→H**	**H→U**	**Avg**
Source Only	73.9	71.7	72.8
TA3N [1]	78.3	81.8	80.1
MA2LT-D [2]	85.0	86.6	85.8
TransferAttn [3]	88.1	88.3	88.2
Ours (ResNet101)	**90.2**	**89.5**	**89.85**

Observation: The proposed method achieves state-of-the-art performance on both transfer directions, outperforming prior approaches by a clear margin. This demonstrates the effectiveness of our approach in learning transferable representations across complex video domain shifts.

Quantitative Results

HHAR Results (Accuracy %)

Alg.	**0→6**	**1→6**	**2→7**	**3→8**	**4→5**	**5→0**	**6→1**	**7→4**	**8→3**	**0→2**	**Avg**
NoAdapt	37.3	56.9	45.1	63.0	63.1	47.5	73.3	63.5	55.4	51.8	55.7
DeepCoral [3]	37.7	56.1	54.5	63.2	65.0	35.2	73.3	70.7	69.8	50.7	57.6
MMDA [4]	38.3	55.3	57.2	53.0	57.3	47.3	**80.2**	76.0	61.1	61.3	58.7
CoDATS [5]	**41.7**	64.3	62.0	75.8	63.2	35.9	59.1	80.0	72.9	58.7	61.4
SASA [6]	44.1	52.9	57.6	69.0	71.8	37.9	72.8	63.7	70.5	62.5	60.3
RAINCOAT [7]	34.6	62.6	65.3	61.1	49.5	37.5	75.1	80.0	62.4	58.0	58.6
SoftMax [2]	38.3	51.9	59.9	70.2	76.8	45.7	79.7	67.9	73.1	61.3	62.5
NCP [9]	35.7	46.1	59.1	42.1	51.8	54.5	52.2	60.9	57.1	55.2	51.5
SP [9]	30.5	44.7	58.5	44.8	57.4	48.1	47.2	55.5	59.7	61.5	50.8
ATT [8]	36.3	52.1	51.6	**77.4**	68.3	47.5	63.6	76.0	59.5	59.6	59.2
SHOT [10]	35.9	47.5	58.2	77.6	**86.5**	44.0	**81.7**	73.9	70.7	**71.6**	64.8
T2PL [11]	37.9	64.7	55.3	73.3	**89.4**	37.9	75.2	77.4	63.7	66.5	64.1
TransPL [12]	39.5	73.1	60.5	72.7	75.4	52.3	77.1	**89.2**	80.3	64.2	68.4
Ours	41.2	**76.5**	**64.2**	76.6	78.2	**58.5**	79.3	88.4	**84.2**	66.7	**71.38**

Evaluation and Results

Ablation: Effect of Causal Constraint and Uncertainty-aware Pseudo-Labeling

Method	**UCIHAR**	**WISDM**	**U $\rightarrow$ H**
Full Causal-OT	**73.97**	**68.03**	**90.2**
w/o Causal Constraint	69.42	63.81	86.4
w/o Uncertainty-aware PL	71.6	65.5	87.9

Observation: Removing either the *causal constraint* or *uncertainty-aware pseudo-labeling* leads to a consistent drop in performance across all datasets. This demonstrates that both components are complementary and jointly contribute to the robustness and effectiveness of the proposed **Causal-OT** framework.

Evaluation and Results

Ablation: Effect of Hyperparameters β and λ (WISDM)

Values	35→31	7→18	17→23	6→19
$\beta = 1$	**72.28**	**81.13**	66.66	**61.36**
$\beta = 2$	69.88	80.19	**68.33**	60.61
$\beta = 3$	69.88	79.49	**68.33**	61.21
$\lambda = 2$	68.67	**81.13**	**68.33**	60.61
$\lambda = 3$	**72.29**	**81.13**	66.67	**61.36**

Observation: The performance remains relatively stable across different values of β and λ, indicating robustness of the proposed method. Slight improvements are observed around $\beta = 1$ and $\lambda = 3$, suggesting these as favorable choices for balancing alignment and regularization.

Effect of Uncertainty Modeling

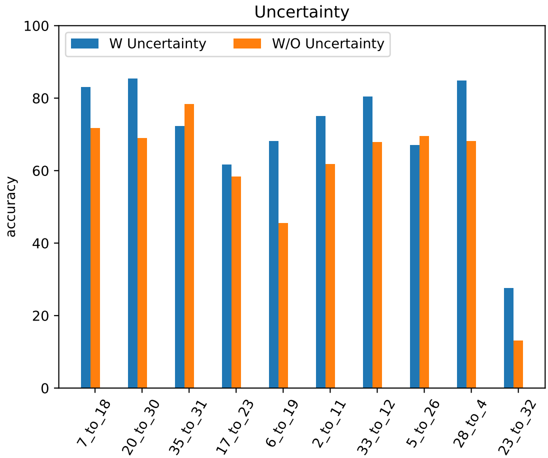

Figure: Domain adaptation performance with and without uncertainty modeling across various source-target domain pairs in the WISDM dataset.

Publication

CVPR2026

Khushboo Mishra, Varun Trivedi, and Tanima Dutta "Towards Uncertainty-aware Unsupervised Domain Adaptation for Videos and Time-Series with Causal Optimal Transport," CVPR Main Track, 2026 (Core A*)(Accepted)

References I

[1] M.-H. Chen, Z. Kira, G. AlRegib, J. Yoo, R. Chen, and J. Zheng, "Temporal attentive alignment for large-scale video domain adaptation," in *Proc. IEEE/CVF ICCV*, pp. 6321–6330, 2019.

[2] P. Chen, Y. Gao, and A. J. Ma, "Multi-level attentive adversarial learning with temporal dilation for unsupervised video domain adaptation," in *Proc. IEEE/CVF WACV*, pp. 1259–1268, 2022.

[3] A. Sacilotti, S. F. Dos Santos, N. Sebe, and J. Almeida, "Transferable-guided attention is all you need for video domain adaptation," in *Proc. IEEE/CVF WACV*, pp. 8691–8701, 2025.

[4] M. M. Rahman, C. Fookes, M. Baktashmotlagh, and S. Sridharan, "On minimum discrepancy estimation for deep domain adaptation," in *Domain Adaptation for Visual Understanding*, pp. 81–94, Springer, 2020.

[5] G. Wilson, J. R. Doppa, and D. J. Cook, "Multi-source deep domain adaptation with weak supervision for time-series sensor data," in *Proc. ACM SIGKDD*, pp. 1768–1778, 2020.

[6] R. Cai, J. Chen, Z. Li, W. Chen, K. Zhang, J. Ye, Z. Li, X. Yang, and Z. Zhang, "Time series domain adaptation via sparse associative structure alignment," in *Proc. AAAI*, vol. 35, no. 8, pp. 6859–6867, 2021.

[7] Q.-Q. He, S. W. I. Siu, and Y.-W. Si, "Attentive recurrent adversarial domain adaptation with Top-k pseudo-labeling for time series classification," *Applied Intelligence*, vol. 53, no. 11, pp. 13110–13129, 2023.

[8] K. Saito, Y. Ushiku, and T. Harada, "Asymmetric tri-training for unsupervised domain adaptation," in *Proc. ICML*, pp. 2988–2997, 2017.

[9] Q. Wang and T. Breckon, "Unsupervised domain adaptation via structured prediction based selective pseudo-labeling," in *Proc. AAAI*, vol. 34, no. 04, pp. 6243–6250, 2020.

[10] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation," in *Proc. ICML*, pp. 6028–6039, 2020.

[11] C. Liu, L. Wang, L. Lyu, C. Sun, X. Wang, and Q. Zhu, "Deja vu: Continual model generalization for unseen domains," *arXiv preprint arXiv:2301.10418*, 2023.

[12] J. Kim and S. Lee, "TransPL: VQ-Code transition matrices for pseudo-labeling of time series unsupervised domain adaptation," *arXiv preprint arXiv:2505.09955*, 2025.

Thank You