

DADA: Dual-Alignment Domain Adaptation for Cross-Scene Pedestrian Trajectory Prediction

Wenzhan Li¹, Fuhao Li¹, Xinghui Jing¹, Pingfa Feng^{1,2}, and Long Zeng¹

Abstract—Cross-scene pedestrian behaviour prediction remains challenging because models trained in one environment often face substantial shifts in motion statistics and scene-dependent behavioural cues when deployed in another. In our recent work [1], we addressed this problem with *Dual-Alignment Domain Adaptation* (DADA), a two-stage framework combining data-level alignment (DLA) and feature-level alignment (FLA). DLA reduces coarse distribution mismatch by translating source trajectories toward target-style motion patterns, while FLA suppresses residual scene-specific bias by aligning latent representations inside the predictor. Cross-domain experiments on ETH&UCY show that DADA consistently improves transfer performance, with gains of up to 53.5% over non-adaptive predictors and 41.7% over prior domain-adaptation methods. This extended abstract revisits these results from a robotics perspective, highlighting behaviour-aware transfer, uncertainty, and deployment-oriented discussion.

I. MOTIVATION AND POSITIONING

Pedestrian behaviour prediction is central to autonomous driving, service robotics, socially compliant navigation, and human-aware planning. In practice, however, a forecaster is often trained on one set of scenes and deployed in another, where crowd density, pedestrian speed, interaction frequency, obstacle layout, and route preference may all change. This mismatch is especially damaging in safety-critical settings, because even a strong in-domain predictor can become brittle once it encounters unfamiliar scene statistics.

Most existing trajectory predictors adopt a multi-scene training paradigm and implicitly expect a unified motion law to generalize everywhere. More recent domain-adaptation approaches such as T-GNN, AML, and BMP improve transferability, but they usually align at a *single level*: the input distribution, latent features, or the output stage [2]–[4]. Our recently published work argues that such a treatment is incomplete, because the transfer gap itself is multi-scale [1].

Specifically, we decompose cross-scene discrepancy into two coupled parts. **Macro differences** capture scene-level motion statistics, including pedestrian density, velocity/acceleration, and interaction patterns. **Micro differences** capture latent scene-dependent behavioural cues, such as local obstacle avoidance, preferred walking styles, and encoded intent. This view motivates a two-stage alignment pipeline: first reducing coarse mismatch at the data level, then correcting residual discrepancy in feature space.

¹Department of Advanced Manufacturing, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. Emails: liwenzhan00@gmail.com, lf23@mails.tsinghua.edu.cn, xhjing99@gmail.com, fengpf@tsinghua.edu.cn, zenglong@sz.tsinghua.edu.cn. ²Department of Mechanical Engineering, Tsinghua University, Beijing, China.

II. DUAL-ALIGNMENT FORMULATION

The overall framework is illustrated in Fig. 1. Let (X_S, Y_S) denote labeled source-domain trajectories and X_T unlabeled target-domain observations. DADA contains two complementary components.

Data-level alignment (DLA). We use two trajectory generators, G_S and G_T , together with two discriminators, D_T and D_S , to translate trajectories between domains. Unlike image translation, trajectory transfer must preserve temporal coherence and socially plausible interactions. We therefore build the generator from temporal convolutional networks (TCNs) and insert a graph-attention layer to retain interaction structure while translating source samples toward the target style. The DLA objective is

$$\begin{aligned} \mathcal{L}_{\text{data}} = & \mathcal{L}_{\text{GAN}}(G_S, D_T) + \mathcal{L}_{\text{GAN}}(G_T, D_S) \\ & + \alpha \mathcal{L}_{\text{cyc}}(G_S, G_T), \end{aligned} \quad (1)$$

where the cycle-consistency term stabilizes translation and limits degenerate motion patterns. The aligned source samples are then mixed with the original source trajectories to form an augmented set X'_S , which serves as the bridge between DLA and the downstream predictor.

Feature-level alignment (FLA). Even after DLA, the encoder may still extract domain-specific cues that hurt generalization. We therefore place an auxiliary discriminator D_E between the encoder and decoder of a standard encoder-decoder forecaster. Given

$$h_S = \text{Enc}(X'_S; \mathbf{W}_{\text{enc}}), \quad h_T = \text{Enc}(X_T; \mathbf{W}_{\text{enc}}), \quad (2)$$

FLA jointly optimizes the prediction loss and an adversarial feature-alignment term,

$$\begin{aligned} \mathcal{L}_{\text{feat}}(\text{Enc}, \text{Dec}, D_E) = & \mathcal{L}_{\text{pred}}(\text{Enc}, \text{Dec}) \\ & + \beta \mathcal{L}_{\text{adv}}(\text{Enc}, D_E). \end{aligned} \quad (3)$$

This design is modular and can be attached to public encoder-decoder predictors without redesigning the forecasting head. DLA addresses observable *macro* shift, while FLA targets residual *micro* shift in the latent representation.

III. REPRESENTATIVE RESULTS AND TAKEAWAYS

We evaluate on ETH&UCY using the 20 cross-domain transfer tasks defined in the original study [1]. As summarized in Table I, DADA improves cross-scene transfer consistently across prior adaptation baselines and diverse forecasting backbones. Against dedicated trajectory domain-adaptation methods [2]–[4], DADA-Trajctron++ and DADA-TUTR reduce average FDE to 0.61 and 0.42,

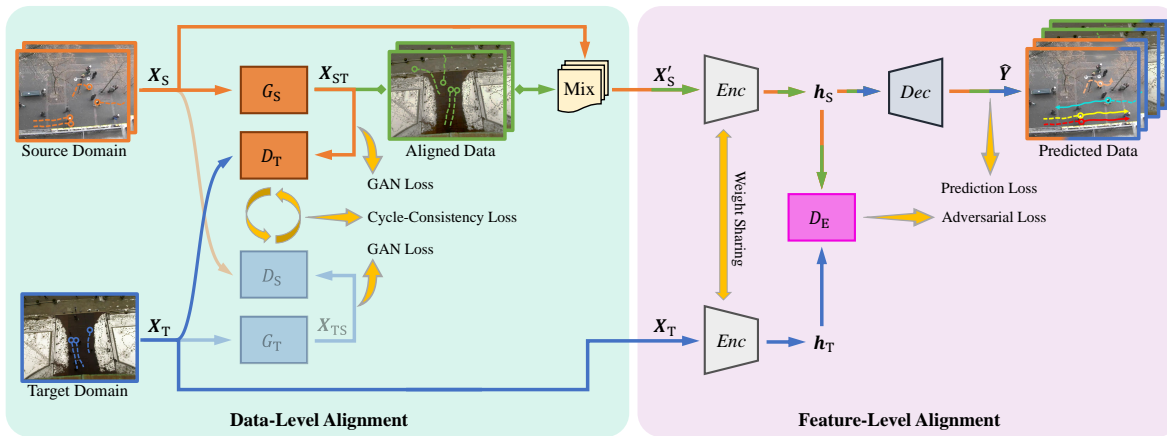


Fig. 1. Overview of the DADA pipeline. As shown in the figure, DLA first translates source trajectories toward target style and mixes the translated samples with the original source data, while FLA then aligns encoder features through an embedded discriminator.

TABLE I
REPRESENTATIVE CROSS-DOMAIN FDE RESULTS, REORGANIZED FROM [1]. LOWER IS BETTER.

Method	Avg. FDE ↓	Comment
T-GNN [2]	1.82	prior DA baseline
AML [3]	1.06	prior DA baseline
BMP [4]	0.72	prior DA baseline
DADA-Trajectron++	0.61	ours
DADA-TUTR	0.42	ours (best)
S-STGCNN → DADA	2.30 → 1.07	53.5% gain
PECNet → DADA	2.29 → 1.15	50.0% gain
Trajectron++ → DADA	1.10 → 0.61	44.5% gain

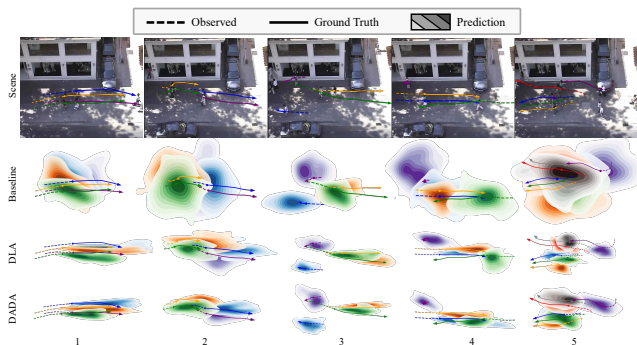


Fig. 2. Qualitative cases based on [1]. Compared with the baseline, DLA corrects coarse directional mismatch, while DADA further restores socially plausible interaction-aware behaviours.

respectively. Two takeaways stand out: 1) hierarchical alignment matters: coarse transfer in data space alone is not enough, but neither is a purely feature-space correction; and 2) the gains are not restricted to one predictor family, suggesting that the proposed decomposition captures a more general transfer principle for pedestrian forecasting.

As shown in Fig. 2, the improvement extends beyond mere numerical performance. In following-heavy scenes, DLA alone recovers the correct dominant walking direction, while the full DADA model further tightens the trajectory distribution around the ground truth. In collision-heavy or turning scenes, DADA also preserves more plausible avoid-

ance and interaction patterns, demonstrating that feature-level alignment complements coarse trajectory translation.

IV. IMPLICATIONS FOR WORKSHOP

These results motivate three WPBP workshop discussion directions. **Behaviour-aware transfer beyond accuracy:** Cross-scene prediction should be judged not only by ADE/FDE, but also by whether following, yielding, and avoidance behaviours persist under deployment shift. **Uncertainty under domain shift:** DADA improves transfer accuracy, yet robotics systems additionally require calibrated multimodal uncertainty, motivating aligning latent features and uncertainty-bearing predictive structure. **Robot-facing deployment:** Since pedestrian forecasting usually feeds a planner or interaction policy, transfer methods should further be evaluated in downstream closed-loop decision making.

V. CONCLUSION

This extended abstract revisits our published DADA results [1] from a workshop perspective. The main message is that cross-scene pedestrian trajectory prediction benefits from hierarchical rather than monolithic adaptation. By separating trajectory-space and latent-space correction, DADA improves transfer accuracy and better preserves behaviourally plausible forecasts. We hope this perspective stimulates discussion on uncertainty-aware transfer and planner-facing evaluation in pedestrian behaviour prediction. Code is available at <https://github.com/jackyzengl/DADA>.

REFERENCES

- [1] W. Li, F. Li, X. Jing, P. Feng, and L. Zeng, "Dual-alignment domain adaptation for pedestrian trajectory prediction," *IEEE Robotics and Automation Letters*, vol. 9, no. 12, pp. 10 962–10 969, 2024.
- [2] Y. Xu, L. Wang, Y. Wang, and Y. Fu, "Adaptive trajectory prediction via transferable gnn," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6520–6531.
- [3] B. Ivanovic, J. Harrison, and M. Pavone, "Expanding the deployment envelope of behavior prediction via adaptive meta-learning," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 7786–7793.
- [4] R. Huang, A. Tompkins, M. Pagnucco, and Y. Song, "Towards single source domain generalisation in trajectory prediction: A motion prior based approach," in *Conference on Lifelong Learning Agents*. PMLR, 2023, pp. 227–243.