Adaptive Trajectory Prediction via Transferable GNN
Supplementary Material

Yi Xu[1], Lichen Wang[1], Yizhou Wang[1], Yun Fu[1,2]
[1]Department of Electrical and Computer Engineering, Northeastern University, USA
[2]Khoury College of Computer Science, Northeastern University, USA
xu.yi@northeastern.edu, {wanglichenxj,wyzjack990122}@gmail.com, yunfu@ece.neu.edu

1. Overview

In the supplementary material, we provide experimental details and more evaluation results including visualizations. We also provide our insights and discussions at the end.

2. Experiments

2.1. Dataset Details

There are two commonly used datasets including five scenes in our work: ETH[1] [8] and UCY[2] [4]. Dataset ETH consists of two scenes: ETH and HOTEL. Dataset UCY consists of three scenes: UNIV, ZARA1, and ZARA2. Each scene contains multiple walking pedestrians with different complex walking motions. We show examples of each scene in Fig. 1 with one red dot as one person. Some basic information of two datasets is shown in Tab. 1, additional statistics of five scenes have been provided in the main body.

2.2. Implementation Details

We compare our proposed model with total 5 trajectory prediction baselines: Social-STGCNN [7], PECNet [6], RSBG [11], SGCN [9], and Tra2Tra [13]. We implemented these baselines with their provided codes: Social-STGCNN [3], PECNet [4], SGCN [5]. We tried our best to reproduce the codes of RSBG and authors have shared the codes of Tra2Tra with us.We also employ 4 domain adaptation approaches: T-GNN+MMD [5] , T-GNN+CORAL [10], T-GNN+GFK [2], and T-GNN+UDA [12]. We implemented these approaches based on the codes: MMD [6], CORAL[7], GFK [8], UDA [9].

2.3. Performance Study of $a_{t;i,j}$

As mentioned in the main body, the value of $a_{t;i,j}$ in adjacency matrix A_t is initialized as the distance between pedestrian i and j,

$$a_{t;i,j} = \|o'^{i}_{t} - o'^{j}_{t}\|_2, \tag{1}$$

where $\| * \|_2$ is the l_2 distance, and o'^{i}_{t} denotes the "relative coordinates" (x'^{i}_{t}, y'^{i}_{t}) of pedestrian i at time step t.

As it should be other possible definitions of $a_{t;i,j}$, thus we investigate and analysis different definitions of $a_{t;i,j}$ as follows. Among these different functions, the key starting point we follow is that $a_{t;i,j}$ could be the function of the relative coordinates of pedestrians i and j. Average ADE and FDE results are shown in Tab. 2.

$$a^{reci}_{t;i,j} = \frac{1}{\|o'^{i}_{t} - o'^{j}_{t}\|_2 + \epsilon}, \tag{2}$$

$$a^{exp}_{t;i,j} = \exp\left(-\frac{\|o'^{i}_{t} - o'^{j}_{t}\|_2^2}{2\sigma^2}\right), \tag{3}$$

$$a^{rat}_{t;i,j} = 1 - \frac{\|o'^{i}_{t} - o'^{j}_{t}\|_2^2}{\|o'^{i}_{t} - o'^{j}_{t}\|_2^2 + c}. \tag{4}$$

where ϵ and c are both two small positive constants to ensure the numerical stability. In real practice, it is really difficult to have $\|o'^{i}_{t} - o'^{j}_{t}\|_2 = 0$ and we set $\epsilon = c = 0.001$ though.

We can see from Tab. 2, the best performance comes from $a_{t;i,j}$ (Eq. (1)). Function $a^{reci}_{t;i,j}$ achieves the second best performance on ADE metric and $a^{rat}_{t;i,j}$ achieves the second best performance on FDE metric, respectively.

[1]http://www.vision.ee.ethz.ch/en/datasets/.
[2]https://graphics.cs.ucy.ac.cy/research/downloads/crowd-data.
[3]https://github.com/abduallahmohamed/Social-STGCNN.
[4]https://github.com/HarshayuGirase/Human-Path-Prediction.
[5]https://github.com/shuaishiliu/SGCN.
[6]https://github.com/easezyc/deep-transfer-learning.
[7]https://github.com/VisionLearningGroup/CORAL.

(a) ETH (b) HOTEL (c) UNIV (d) ZARA1 (e) ZARA2

Figure 1. One frame example from five different scenes. All five scenes are from outdoor top-down view where multiple pedestrians walking in different motions (each person is denoted with one red dot). It is obvious that UNIV is much more crowded than other four scenes. In addition, ZARA1 and ZARA2 share almost the same background.

Dataset	Year	Location	Target	Sensors	Description	Duration and tracks	Annotations	Sampling
ETH	2009	Outdoor	People	Camera/Top-down view	Two scenes	25 min, 650 tracks	Positions, velocities, groups, maps	@2.5Hz
UCY	2007	Outdoor	People	Camera/Top-down view	Three scenes	16.5 min, over 700 tracks	Positions, gaze directions	–

Table 1. Basic information ETH and UCY.

Variants	Performance	
	ADE	FDE
$a_{t;i,j}^{reci}$	$\underline{1.03}$	1.99
$a_{t;i,j}^{exp}$ $(\sigma = 2)$	1.17	2.10
$a_{t;i,j}^{exp}$ $(\sigma = 4)$	1.09	1.99
$a_{t;i,j}^{exp}$ $(\sigma = 8)$	1.14	2.07
$a_{t;i,j}^{rat}$	1.08	$\underline{1.92}$
$\boldsymbol{a_{t;i,j}}$	**0.96**	**1.82**

Table 2. Average performance of total 20 tasks on ADE/FDE metric with different initializations for the adjacency matrix A_t.

2.4. Results of Other DA Approaches

Tab. 3 shows evaluation results of total 20 tasks when comparing with other domain adaptation approaches. For model T-GNN+UDA, in which there is an adversarial loss that measured by an extra domain classifier. One fully-connected linear layer is employed as the classifier. In specific, this kind of models needs to minimize the adversarial loss with respect to parameters of domain classifier, while maximizing it with respect to the parameters of trajectory predictor. Thus we use a a gradient reversal layer [1] for the min-max optimization to unify the training procedure in a single step. It can be observed that our proposed adaptive learning module outperforms these domain adaptation approaches. This may show that our designed alignment loss is more appropriate for adapting fine-grained individual-level features in trajectory prediction task.

[8]https://github.com/jindongwang/transferlearning/tree/master/code/traditional/GFK.

[9]https://github.com/GRAND-Lab/UDAGCN.

2.5. t-SNE Visualization

In this section, we visualize feature representations $F_{(s)}$ and $F_{(t)}$ of the target and source trajectory domain with t-SNE [3] approach on of 4 tasks. Fig. 2 shows the visualization examples where red and blue denote the source and target trajectory features, respectively. The first row are $F_{(s)}$ and $F_{(t)}$ without attention-based adaptive learning module , which we denote as "w/o AAL" (corresponding to the Variant T-GNN-V_2 in the main body). The second row are with attention-based adaptive learning module, which we denote as "w AAL". Each dot represents the feature of one pedestrian in the figure. Different from conventional t-SNE visualizations that applied in classification task, there is **no specific "label"** of each dot in our task. Therefore, the cluster structure may not be clear in our task.

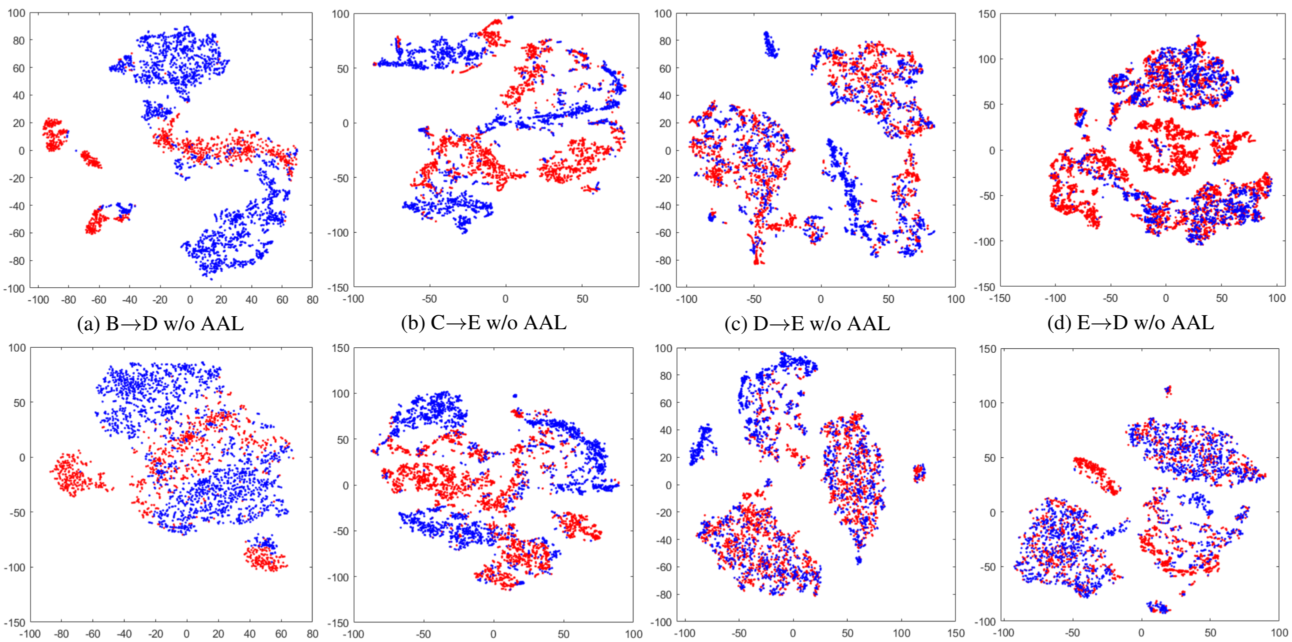
For task B→D and C→E, we can observe that features get closer with our adaptive learning module, which validates that our proposed adaptation learning module is able to alleviate the disparities across different trajectory domains. In addition, the visualization of task B→D is not significant and the corresponding quantitative results of task is B→D lower than others (ADE: 2.25, FDE:4.04). For task D→E and E→D, since D (ZARA1) and E (ZARA2) have similar scenes, we can observe from these two pairs of figures: (1) features from source and target domains have more overlaps, (2) features become more closer. It is consistent with their corresponding quantitative results (D→E: 0.32/0.65, E→D: 0.34/0.72). It also validates the effectiveness of our proposed adaptive learning module.

3. Discussion

Compare with general domain adaptation methods. We delve into the domain-shift challenge in the task of pedestrian trajectory prediction in this paper. In image/video-

Metric	Method	Performance (Source2Target)																				Ave
		A2B	A2C	A2D	A2E	B2A	B2C	B2D	B2E	C2A	C2B	C2D	C2E	D2A	D2B	D2C	D2E	E2A	E2B	E2C	E2D	
ADE	T-GNN+MMD [5]	1.53	1.39	1.14	1.19	2.99	1.18	2.39	1.49	1.08	0.62	0.71	0.42	1.02	0.89	0.68	0.38	0.89	0.99	0.74	0.41	1.11
	T-GNN+CORAL [14]	1.43	1.35	1.09	1.12	2.87	1.12	2.31	1.46	1.03	0.58	0.68	0.46	0.99	0.85	0.66	0.40	0.86	0.96	0.67	0.41	1.07
	T-GNN+GFK [2]	1.69	1.52	1.20	1.24	3.01	1.19	2.52	1.55	1.11	0.68	0.69	0.50	0.96	0.89	0.71	0.43	0.89	1.01	0.75	0.42	1.15
	T-GNN+UDA [12]	1.41	1.32	0.98	1.23	2.92	1.20	2.43	1.42	1.12	0.64	0.62	0.48	0.91	0.81	0.69	0.35	0.91	0.98	0.70	0.39	1.07
	T-GNN (Ours)	**1.13**	**1.25**	**0.94**	**1.03**	**2.54**	**1.08**	**2.25**	**1.41**	**0.97**	**0.54**	**0.61**	**0.23**	**0.88**	**0.78**	**0.59**	**0.32**	**0.87**	**0.72**	**0.65**	**0.34**	**0.96**
FDE	T-GNN+MMD [5]	2.63	2.65	1.98	2.24	4.86	2.15	4.63	2.69	2.16	1.25	1.52	0.99	2.20	1.88	1.39	0.75	2.03	1.84	1.46	0.82	2.11
	T-GNN+CORAL [14]	2.44	2.52	1.82	2.16	4.59	1.89	4.48	2.68	2.09	1.20	1.47	0.97	2.09	1.83	1.33	0.75	2.01	1.79	1.38	0.79	2.01
	T-GNN+GFK [2]	2.67	2.66	2.03	2.21	4.74	2.12	4.88	2.68	2.19	1.23	1.34	1.01	1.96	1.77	1.30	0.76	2.03	1.83	1.43	0.78	2.08
	T-GNN+UDA [12]	2.59	2.61	1.94	2.25	4.81	2.13	4.85	2.63	2.19	1.29	1.42	1.03	2.03	1.75	1.37	0.73	2.08	1.80	1.45	0.76	2.09
	T-GNN (Ours)	**2.18**	**2.25**	**1.78**	**1.84**	**4.15**	**1.82**	**4.04**	**2.53**	**1.91**	**1.12**	**1.30**	**0.87**	**1.92**	**1.46**	**1.25**	**0.65**	**1.86**	**1.45**	**1.28**	**0.72**	**1.82**

Table 3. ADE/FDE results of our T-GNN model in comparison with existing domain adaptation approaches on 20 tasks. "2" represents from source trajectory domain to target trajectory domain. A, B, C, D, and E denote ETH, HOTEL, UNIV, ZARA1, and ZARA2, respectively.



(a) B→D w/o AAL (b) C→E w/o AAL (c) D→E w/o AAL (d) E→D w/o AAL

(e) B→D w/ AAL (f) C→E w/ AAL (g) D→E w/ AAL (h) E→D w/ AAL

Figure 2. Visualization results of the feature representations $F_{(s)}$ and $F_{(t)}$ using t-SNE. The blue and red dots denote the source and target feature representation, respectively. "w/o AAL" denotes that we disregard attention-based adaptive learning module(corresponding to the Variant T-GNN-V_2). "w/ AAL" denotes $F_{(s)}$ and $F_{(t)}$ are extracted from our proposed T-GNN model.

related classification tasks, domain adaptation (DA) is a hot topic that aims to enable models to generate to novel datasets with different sample distributions. In this study, we expose the challenging domain-shift issue in future trajectory prediction. We usually consider the trajectory as two parts: observation and prediction. It is different from conventional DA tasks where data is in the form of sample-label pairs. Strictly speaking, in trajectory prediction task, the prediction part is not exactly the "label" of the observation part. This essential difference brings in another interesting finding that is worth exploring. In Fig. 2, the cluster structure is not clear because there is no category of each trajectory, which means there exists distribution overlap of different trajectory domains. This kind of "overlap" may be reason of the variance of different tasks. If this "overlap" problem as well as domain-shift problem can be well-addressed simultaneously, trajectory prediction task would be more practical and promising. On the other hand, the observation and prediction parts of one trajectory are totally consistent, thus these two parts may be able to swap and supervise each other. We hope this perspective will inspire the research communities of considering the trajectory prediction problem as well as domain shift issue.

References

[1] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the International Conference on Machine Learning*, pages 1180–1189. PMLR, 2015. 2

[2] Boqing Gong, Yuan Shi, Fei Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2066–2073, 2015. 1, 3

[3] Van Der Maaten Laurens and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2605):2579–2605, 2008. 2

[4] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. *Computer Graphics Forum*, 26(3):655–664, 2010. 1

[5] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *Proceedings of the International Conference on Machine Learning*, pages 97–105, 2015. 1, 3

[6] Karttikeya Mangalam, Harshayu Girase, Shreyas Agarwal, Kuan Hui Lee, Ehsan Adeli, Jitendra Malik, and Adrien Gaidon. It is not the journey but the destination: Endpoint conditioned trajectory prediction. In *Proceedings of the European Conference on Computer Vision*, pages 759–776, 2020. 1

[7] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 14424–14432, 2020. 1

[8] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc J. Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 261–268, 2009. 1

[9] Liushuai Shi, Le Wang, Chengjiang Long, Sanping Zhou, Mo Zhou, Zhenxing Niu, and Gang Hua. Sgcn: Sparse graph convolution network for pedestrian trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8994–9003, 2021. 1

[10] Baochen Sun and Kate Saenko. Deep CORAL: Correlation alignment for deep domain adaptation. In *Proceedings of the European Conference on Computer Vision*, pages 443–450. Springer, 2016. 1

[11] Jianhua Sun, Qinhong Jiang, and Cewu Lu. Recursive social behavior graph for trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 660–669, 2020. 1

[12] Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. Unsupervised domain adaptive graph convolutional networks. In *Proceedings of the Web Conference*, pages 1457–1467, 2020. 1, 3

[13] Yi Xu, Dongchun Ren, Mingxia Li, Yuehai Chen, Mingyu Fan, and Huaxia Xia. Tra2tra: Trajectory-to-trajectory prediction with a global social spatial-temporal attentive neural network. *IEEE Robotics and Automation Letters*, 6(2):1574–1581, 2021. 1

[14] Junbao Zhuo, Shuhui Wang, and Weigang Zhang. Deep unsupervised convolutional domain adaptation. In *Proceedings of the ACM International Conference on Multimedia*, pages 261–269, 2017. 3