Spatial-Temporal Consistency Network for Low-Latency Trajectory Forecasting Supplementary Material

Shijie Li, Yanying Zhou, Jinhui Yi, and Juergen Gall University of Bonn

1. Network Architecture Search

In this section, we provide additional details of the network architecture search. We use the one-shot search algorithm ENAS [3] and implement it based on the open-source NAS toolkit NNI (Neural Network Intelligence)¹. To define a good search space, we use the below settings: 1) ENAS can select the depth of the network; 2) for each layer before the feature-wise convolution, ENAS can choose from temporal convolutions with different dilation rates and graph convolutions; 3) for each layer after feature-wise convolution, ENAS can choose only temporal convolutions with different dilation rates. As for the reward function, we compute the average ADE (average displacement error) of 20 sampled trajectories:

$$R = \beta - Avg(ADE) \tag{1}$$

where Avg(ADE) is the average ADE and β is set as 2 to make the reward positive. We show the architectures of STC-Net and STC-Net-NAS in Tab. 1. The neural network architecture search inverses the order of the dilation rates before the feature-wise convolution and uses only one instead of two graph convolutions. While STC-Net uses the order GC-DTC-GC, STC-Net-NAS uses the order DTC-GC-DTC. After the feature-wise convolution, STC-Net-NAS uses one layer less than STC-Net and the dilation rates are always 2.

2. Impact of Observed Trajectory for Refinement

In Table 3 of the paper, we show that if we do not concatenate the reconstructed observed and the forecast trajectories but refine only the forecast trajectory (w/o obs), the error increases. This is illustrated in Fig. 1. The qualitative results show that including the observed reconstructed trajectory for refinement results in more plausible trajectories. Note that the single yellow dots are standing persons that do not move.

¹Microsoft NNI: https://github.com/microsoft/nni

STC-Net	STC-Net-NAS
GC	DTC(4)
DTC(1)	GC
DTC(2)	DTC(2)
DTC(4)	DTC(1)
GC	
FWC	FWC
DTC(1)	DTC(2)
DTC(2)	DTC(2)
DTC(4)	DTC(2)
DTC(8)	DTC(2)
DTC(16)	

Table 1: Architectures of STC-Net and STC-Net-NAS. GC denotes graph convolution, DTC(D) denotes dilated temporal convolution with dilation rate d, and FWC denotes feature-wise convolution.



Figure 1: Qualitative comparison of forecast trajectories if the reconstructed observed trajectory is included or not for refinement. The **red** line is the observed trajectory, the **blue** line is the ground truth of the future trajectory and the **yellow** dashed line is the prediction.

3. More Qualitative Results

Finally, we show more qualitative results in Fig. 2 and Fig. 3.



Figure 2: Qualitative results. The **red** line is the observed trajectory, the **blue** line is the ground truth of the future trajectory and the **yellow** dashed line is the prediction.



(a) Social-STGCNN [2] (b) STGAT [1] (c) STC-Net Figure 3: Qualitative results. The **red** line is the observed trajectory, the **blue** line is the ground truth of the future trajectory and the **yellow** dashed line is the prediction.

References

- Yingfan Huang, Huikun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In *ICCV*, pages 6272–6281, 2019. 2, 3
- [2] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *CVPR*, pages 14424–14432, 2020. 2, 3
- [3] Hieu Pham, Melody Y Guan, Barret Zoph, Quoc V Le, and Jeff Dean. Efficient neural architecture search via parameter sharing. *arXiv preprint arXiv:1802.03268*, 2018. 1