Appendix for "Traj2Former: A Local Context-aware Snapshot and Sequential Dual Fusion Transformer for Trajectory Classification"

Yuan Xie
National University of
Singapore
yuan_xie@nus.edu.sg

Yichen Zhang
National University of
Singapore
yichenz@nus.edu.sg

Yifang Yin
Institute for Infocomm
Research (I^2R), A*STAR
yiny@i2r.a-star.edu.sg

Sheng Zhang
Institute for Infocomm
Research (I^2R), A*STAR
zhangs2@i2r.a-star.edu.sg

Ying Zhang
Northwestern
Polytechnical University
izhangying@nwpu.edu.cn

Rajiv Ratn Shah
IIIT-Delhi
rajivratn@iiitd.ac.in

Roger Zimmermann
National University of
Singapore
rogerz@nus.edu.sg

Guoqing Xiao
Hunan University
xiaoguoqing@hnu.edu.cn

1 Dataset

Table 1 presents a detailed description of our datasets. The whole trajectory of these two datasets is initially split into discrete trips if the time interval between two consecutive points exceeds 20 minutes. Then, segments will be counted at a fixed length as input for the model.

Geolife. In Geolife, 27,795 segments are extracted from the original trajectories. To fix the length of segments with 600 GPS points, every consecutive point is inserted into one GPS point at every interval. Geolife contains six modes of transportation, and the ratio includes the Walk (32.27%), Bike (12.47%), Car (20.55%), Bus (22.96%), Subway (4.46%), and Train (7.19%). The geographical span of the collected GPS points ranges from 18.24°N to 55.97°N latitude and from 122.33°E to 126.99°W longitude.

MTL. In MTL, 23,406 segments are obtained. Each GPS point is inserted every two consecutive points with 5-second intervals, and the segment length is 650. Four modes of transportation are contained and the ratio is Walk (14.97%), Bike (29.12%), Bus (29.40%), Public (26.51%). The scope of MTL is a geographical spread from $45.23^oN \sim 45.99^oN$ latitude to $72.81^oE{\sim}74.31^oE$ longitude.

Table 1: Dataset Statics of the two datasets.

Dataset	Geolife	MTL
Segments	27,795	23,406
Points	Max: 600	Max: 650
Interpolation	1s	5s
Start Date	2007-04-01	2016-10-17
End Date	2012-08-31	2016-11-17
Class Ratio	Walk:32.37%	Walk:14.97%
	Bike:12.47%	Bike:29.12%
	Car:20.55%	Bus:29.40%
	Bus:22.96%	Public:26.51%
	Subway:4.46%	-
	Train:7.19%	-
MBR	$18.24^oN - 55.97^oN$	$45.23^oN - 45.99^oN$
	$122.33^oE - 126.99^oW$	$72.81^oE - 74.31^oE$

2 Ablation Studies

In this section, we show the experiment results of comparison to both traditional and advanced methods on more datasets, and the impact of different self-Image generation methods.

2.1 UCI and Grab

To present the superior performance of our proposed Traj2Former model, we evaluate over existing methods across different datasets, UCI dataset[1] and Grab-posisi[2]. Table 2 shows our Traj2Former model achieves 97.78% and 88.26% accuracy on the UCI and Grab datasets, respectively.

Table 2: Comparison to Both Traditional and Advanced Methods on More Datasets.

Methods	Acc(UCI)	Acc(Grab-posisi)
Traj2Former	97.78%	88.26%
SECA	67.52%	55.26%
BiLSTM	78.41%	57.19%
TraClets	77.07%	61.23%
Estimator	79.30%	62.25%

2.2 Different self-image generation methods.

We experimented with various methods for self-image generation of trajectory data, including FRFS, FS, and FRRS. The FRFS method involved a fixed range size and pixel size. In this approach, each pixel's coverage range is constant, i.e., 0.0003, which is approximately 33.3 meters in our setting. Therefore, this method generates different pixel sizes to cover the trajectory with different distances. For example, a long-distance trajectory like Train generating an image necessitates a substantially larger pixel size, while shorter trajectories, like Walk, require smaller pixel sizes. This discrepancy arises from the differing speeds of travel and the same time duration for each segment. To standardize the image size across diverse trajectories, FRFS involves cropping a portion of the excessively large image or applying zero-padding to smaller images to maintain a uniform image size.

Table 3: Impact of Different Self-Image Generation Methods.

Pixel Range	Acc	Δ
FRFS	67.82%	-4.07%
FS	69.08%	-2.81%
FRRS, P_r=0.0001	72.00%	-0.84%
FRRS, P_r=0.0003	72.84%	-
FRRS, P_r=0.0005	72.11%	-0.73%
FRRS, P_r=0.0007	72.01%	-0.83%

Since the FRFS method maintains the fixed range within each pixel, the cropping operation reduces the information on the long-distance trajectory. To fully capture a trajectory's details, we introduce the fixed pixel size (FS) method: after computing the maximum latitude and longitude, we adjust the pixel size to a fixed value (e.g., 50 pixels) to generate the self-image. However, this self-image generation technique ignores addressing the scale variations between long-distance and short-distance trajectories, such as Train having a large pixel range of trains while Walk having a small one.

To overcome the limitation of the FRFS and FS methods, we devised an adaptive method named fixed range and reshape size (FRRS). We initially generate self-images in a fixed pixel range to cover the whole trajectory by creating different pixel sizes for different trajectories. Subsequently, to achieve standardized images with uniform pixel sizes, the image is reshaped to a fixed pixel size (i.e., 50x50). Therefore, FRRS not only contains the whole trajectory but also includes the scale variations of different length trajectories.

Results. Table 3 shows the effectiveness of self-image generation methods, and we apply the CNN model to identify the classification and test their effectiveness. The default pixel range P_r is 0.0003, and the pixel size is 50. Our FRRS method outperformed FRFS and FS, with improvements of 4.07% and 2.81%, respectively, demonstrating the importance of including full trajectory information and addressing scale variation. Adjusting the pixel range from 0.0001 to 0.0007, we found that accuracy initially rises as more GPS data is captured from the single pixel. However, further expansion causes a decline in accuracy because most GPS points are located in the same pixels.