NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Generating Mobility Trajectories with Retained Data Utility

Chu Cao, Mo Li
Nanyang Technological University
Singapore

ACM SIGKDD 2021, Virtual Event

1. Background and Motivation

A ***smart city*** is an urban development vision to integrate multiple information and communication technology (ICT) and Internet of Things (IoT) solutions in a secure fashion to manage a city's assets.

——— Wikipedia

http://www.smartnation.sg

1. Background and Motivation

Traffic Surveillance System

Cell Towers

GPS Satellites

Mobility Trajectory Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background and Motivation

Transportation system is a key component in smart cities.

Mobility Trajectory. A mobility trajectory consists of a sequence of locations. The ith trajectory can be denoted as $\tau_i = \{loc_1^i, ..., loc_n^i\}$, where loc_j^i is the jth location sampled at time t_j.

Location. A location is determined by three elements: latitude, longitude, and timestamp, denoted as $loc = (lat, lon, t)$.

Mobility trajectories could help improve the transportation systems.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background and Motivation

GPS readings from vehicles moving in cities could be used to estimate traffic condition [1] and predict non-recurrent traffic events [2].

[1] Z. Liu, et al. Think Like A Graph: Real-Time Traffic Estimation at City-Scale. IEEE TMC 2019.
[2] M. Li, et al. Traffic Flow Prediction via Vehicle Trajectories. AAAI 2021.
https://www.stickpng.com/img/miscellaneous/gps/gps-satellite

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background and Motivation

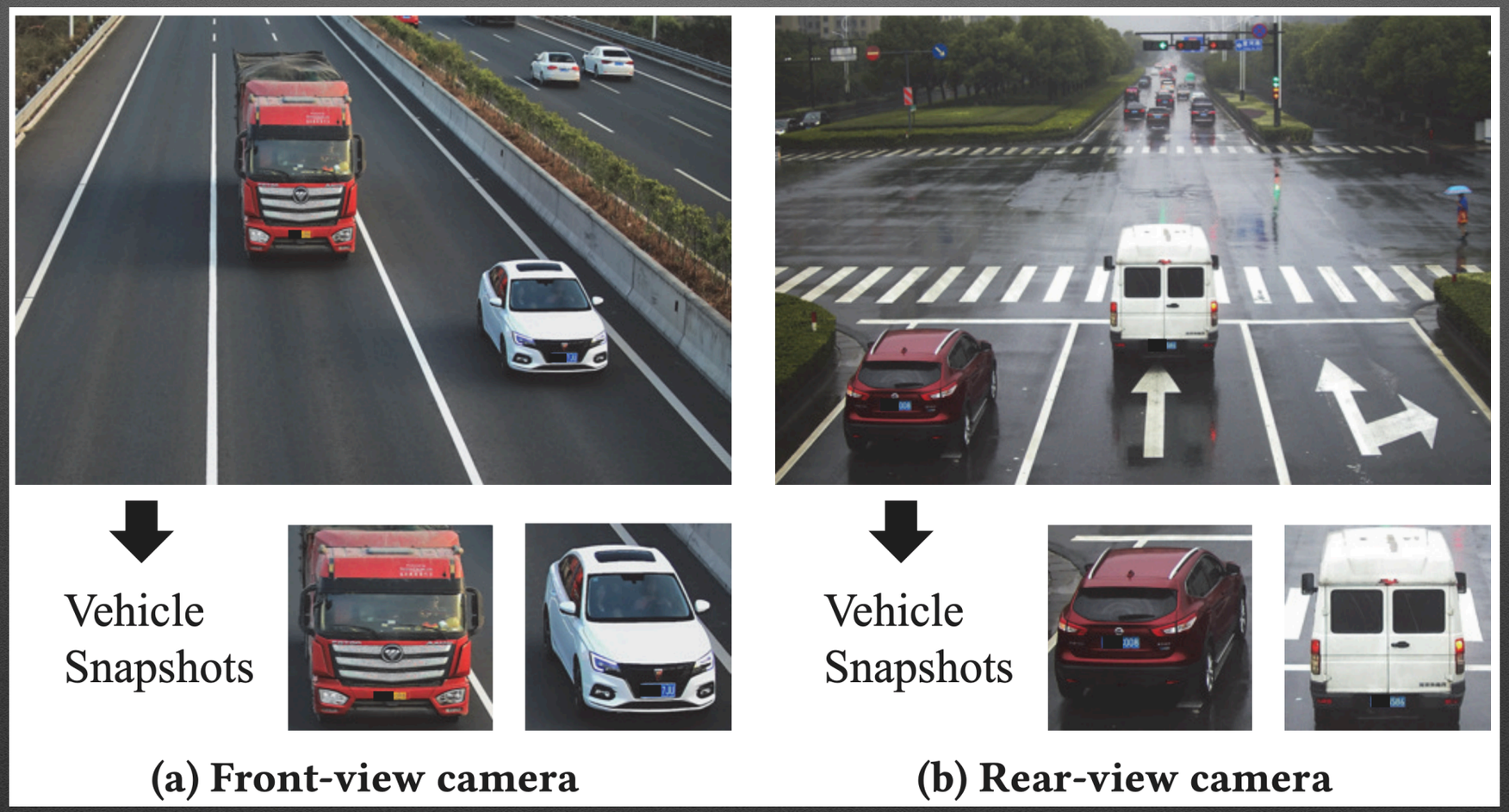
Pedestrian mobilities can be used to derive the uncharted walkways [1].

[1] C. Cao, et al. Walkway Discovery from Large Scale Crwodsensing. In IEEE/ACM IPSN 2018.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background and Motivation

Vehicles' movements captured by the traffic surveillance system help reconstruct the exact trajectories of vehicles in the city [1].



[1] P. Tong, et al. Large-Scale Vehicle Trajectory Reconstruction with Camera Sensing Network. In ACM MobiCom 2021.
https://www.pinclipart.com/downpngs/ioRRhhh_cctv-symbol-cctv-cameras-clipart-png-download/

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background and Motivation

[1] Z. Fang, et al. MoCha: Large-Scale Driving Pattern Characterisation for Usage-based Insurance. In ACM ***SIGKDD*** 2021.

[2] G. Wang, et al. Joint Real-Time Repositioning and Charging for Electric Carsharing with Dynamic Deadlines. In ACM ***SIGKDD*** 2021.

[3] H. Ren, et al. MTrajRec: Map-Constrained Trajectory Recovery via Seq2Seq Multi-task Learning. In ACM ***SIGKDD*** 2021.

[4] Z. Qin, et al. MIMU: Mobile WiFi Usage Inference by Mining Diverse User Behaviours. In ACM ***UbiComp*** 2021.

[5] D. Zhao, et al. D2Park: Diversified Demand-aware On-street Parking Guidance. In ACM ***UbiComp*** 2021.

[6] G. Wang, et al. Data-Driven Fairness-Aware Vehicle Displacement for Large-Scale Electric Taxi Fleets. In IEEE ***ICDE*** 2021.

[7] Y. Yang, et al. VeMo: Enabling Vehicular Mobility Modeling at Individual Levels with Full Penetration. In ACM ***MobiCom*** 2019.

[8] D. Zhang, et al. Exploring Human Mobility with Multi-Source Data at Extremely Large Metropolitan Scales. In ACM ***MobiCom*** 2014.

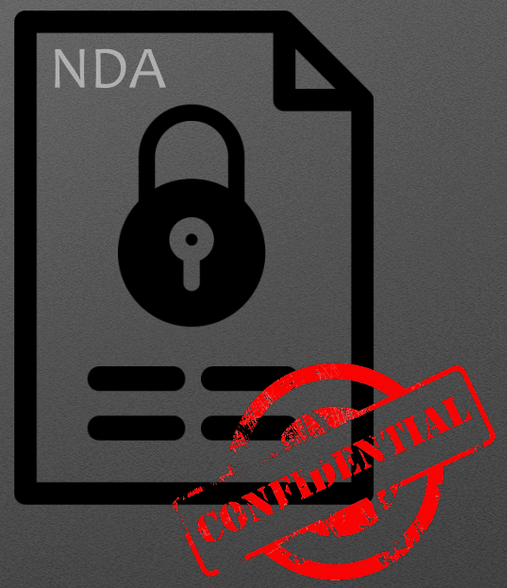
…

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

1. Background and Motivation

Agencies:

- Land Transportation Authority
- EZ-link company
- Alibaba Group
- A-start Research



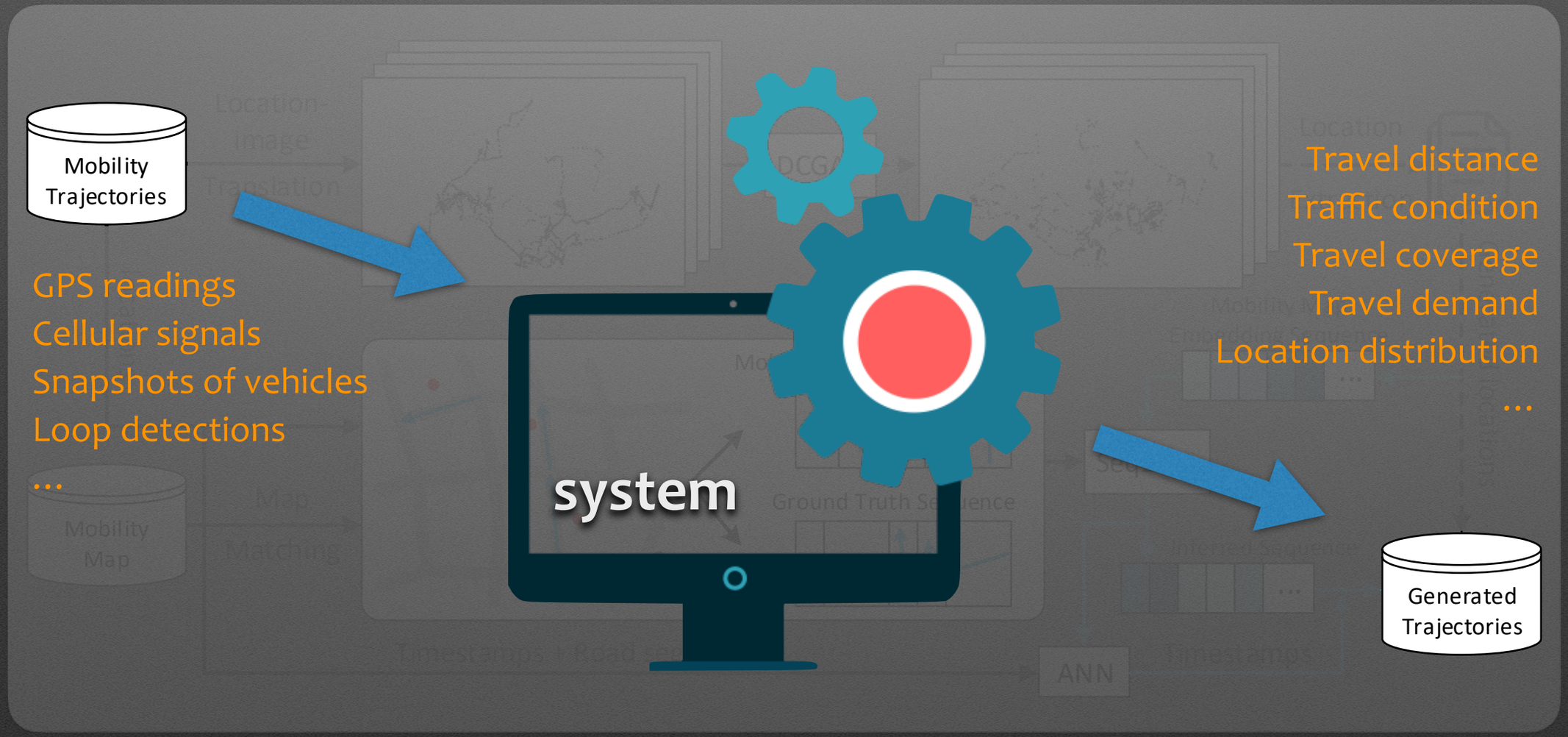
Non-disclosure Agreement documents disable data sharing.

- Comparison on the same dataset
- Validation on the original dataset

https://thenounproject.com/term/nda-non-disclosure-agreement/3302715/
https://www.pikpng.com/downpngs/moxmhx_confidential-data-re-la-96-nike-missile-site/

1. Background and Motivation

Given a set of trajectories $\{\tau_1, \tau_2, ..., \tau_m\}$, and mobility map information, we want to generate a new dataset of mobility trajectories $\{\hat{\tau_1}, \hat{\tau_2}, ..., \hat{\tau_m}\}$, where τ_i and $\hat{\tau_i}$ are the original trajectory and the newly generated trajectory, respectively.

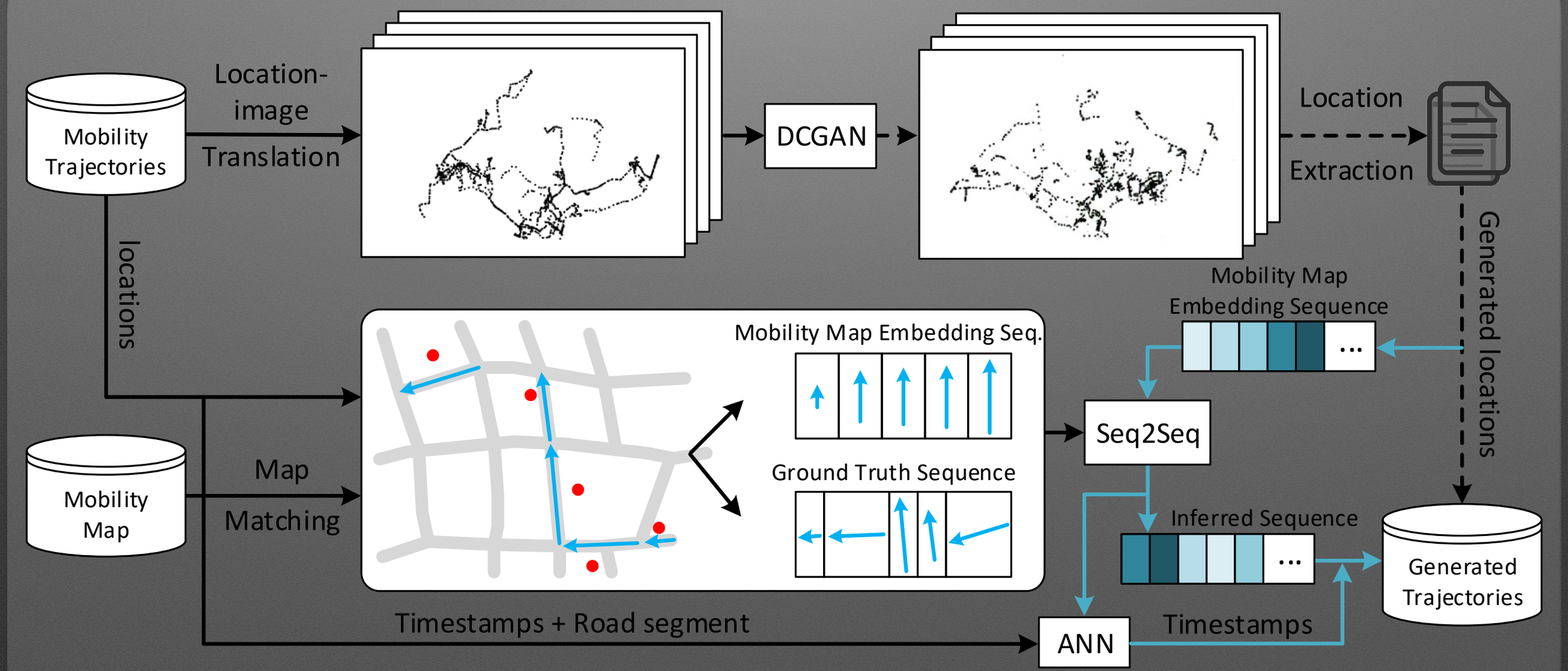


2. Related Work

- Noise + Raw Data
  - Hard to balance utility and privacy
  - Undermine the data utility
- Mixing Raw Data
  - Have potential chance to recover raw data
- ML-based Synthesization
  - Essentially are Noise + Raw Data
  - Only consider spatial information

3. System Design

To generate mobility trajectories with retained data utility.

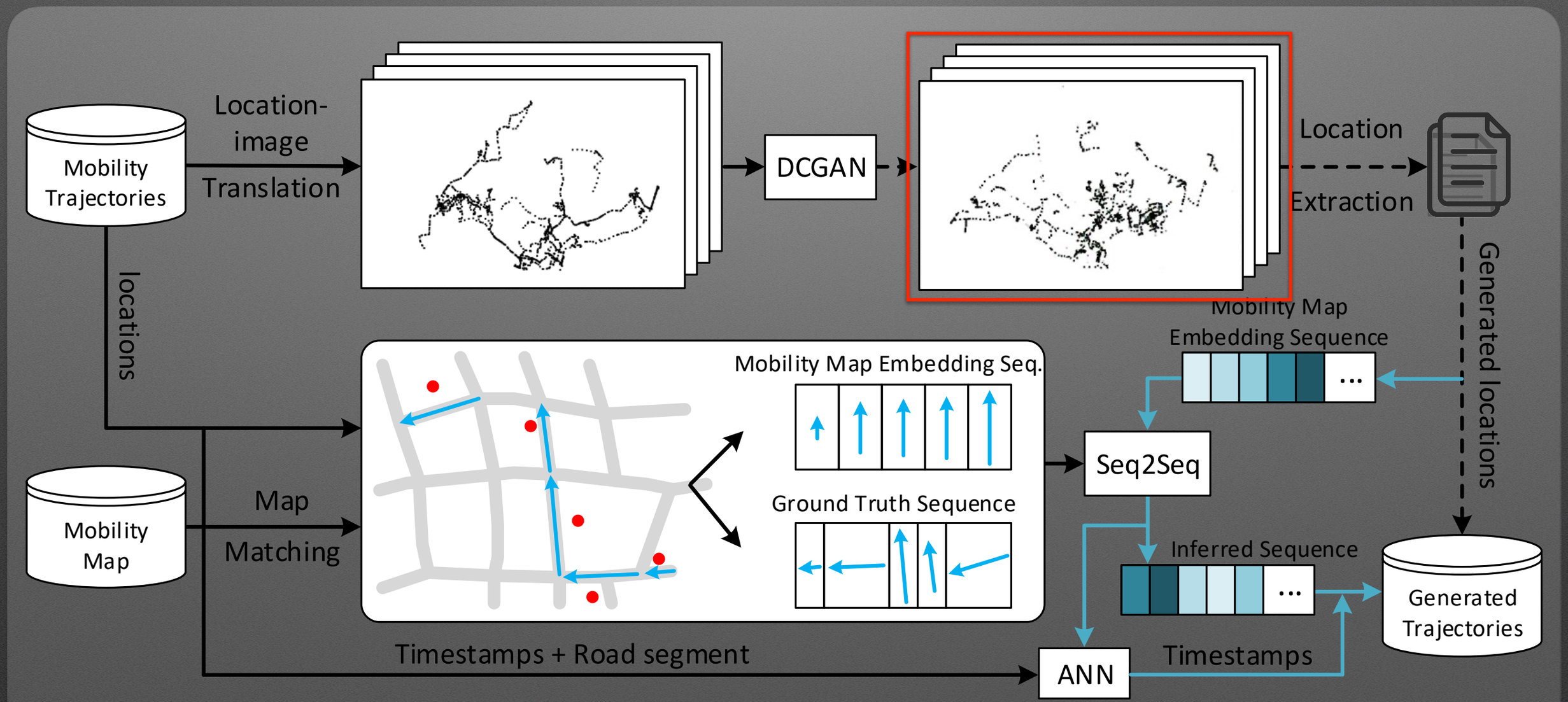


System architecture of TrajGen

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

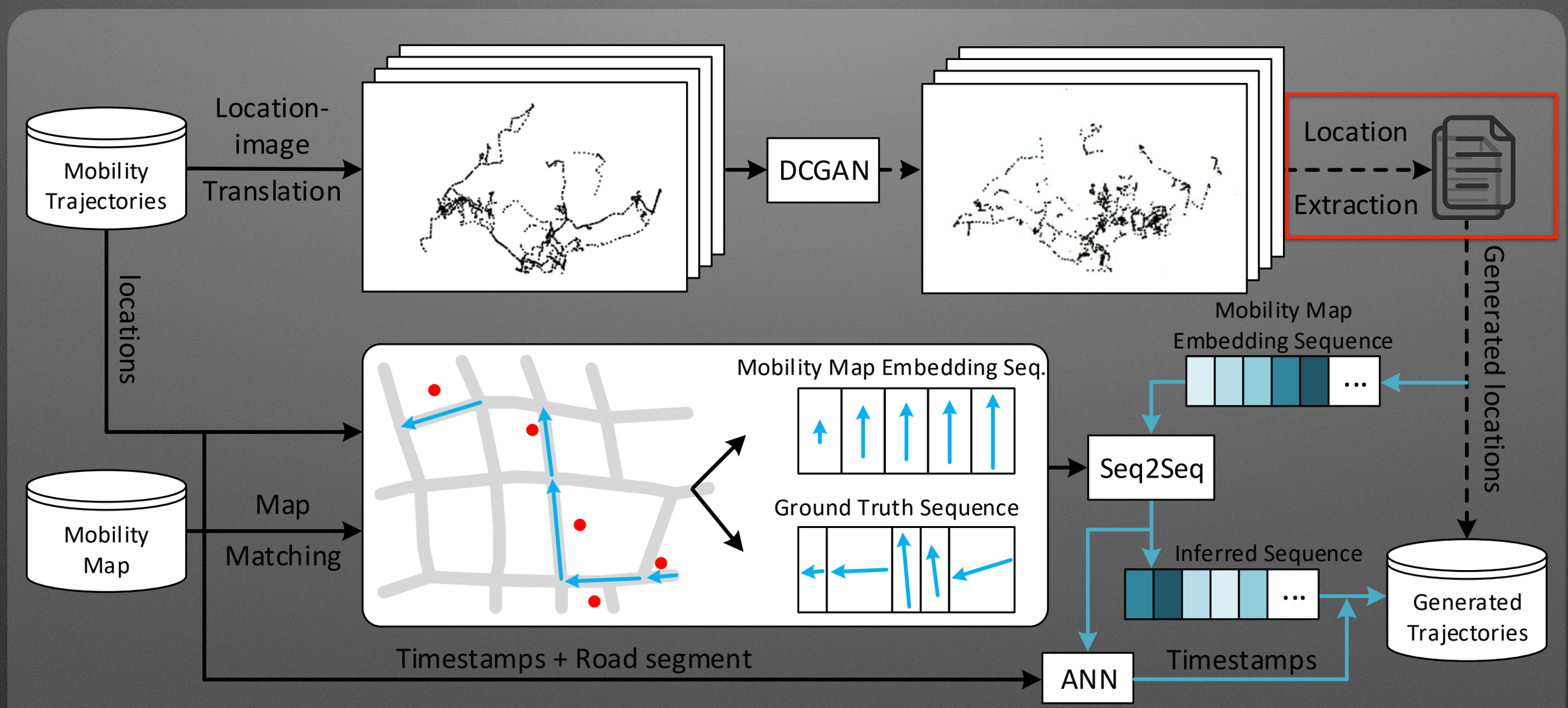
3. System Design

Generate new trajectories in TrajGen.



3. System Design

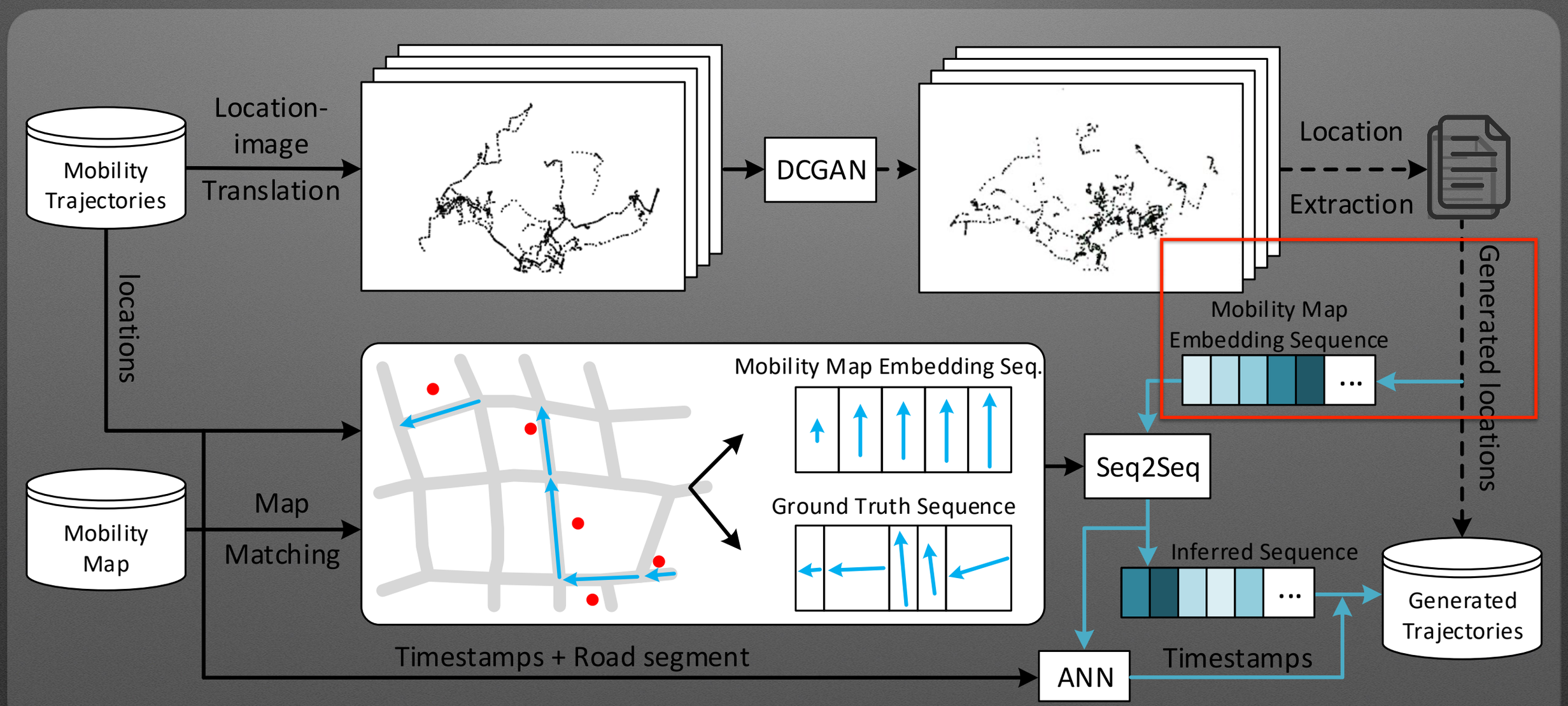
Generate new trajectories in TrajGen.



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

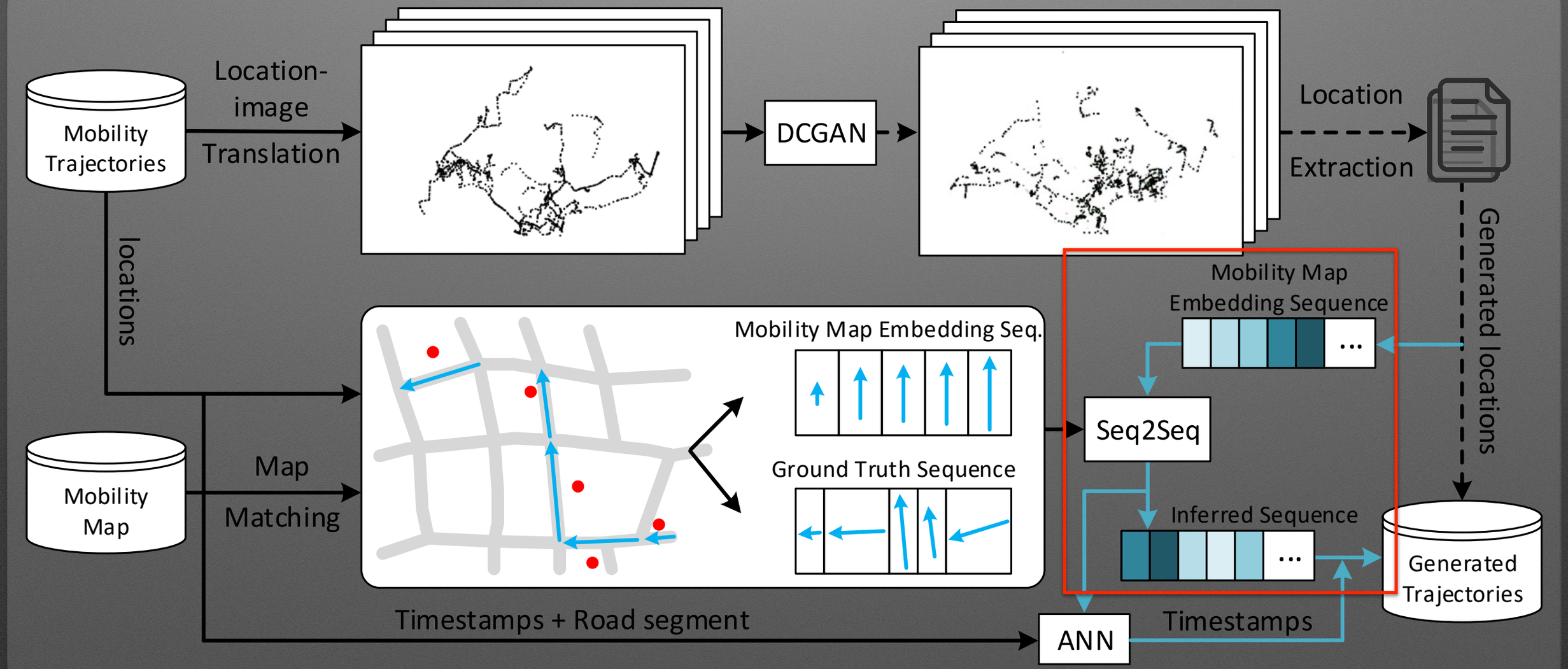
Generate new trajectories in TrajGen.



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

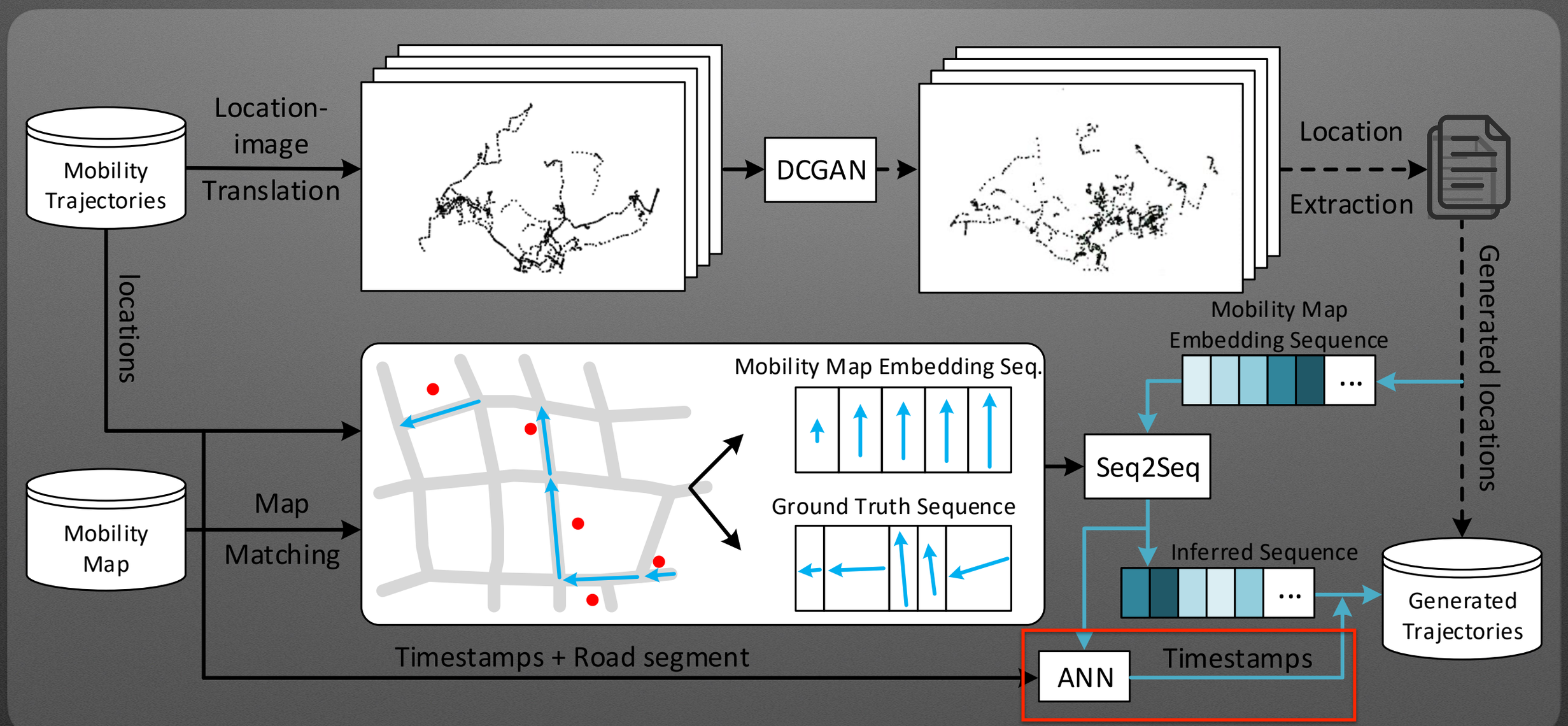
3. System Design

Generate new trajectories in TrajGen.



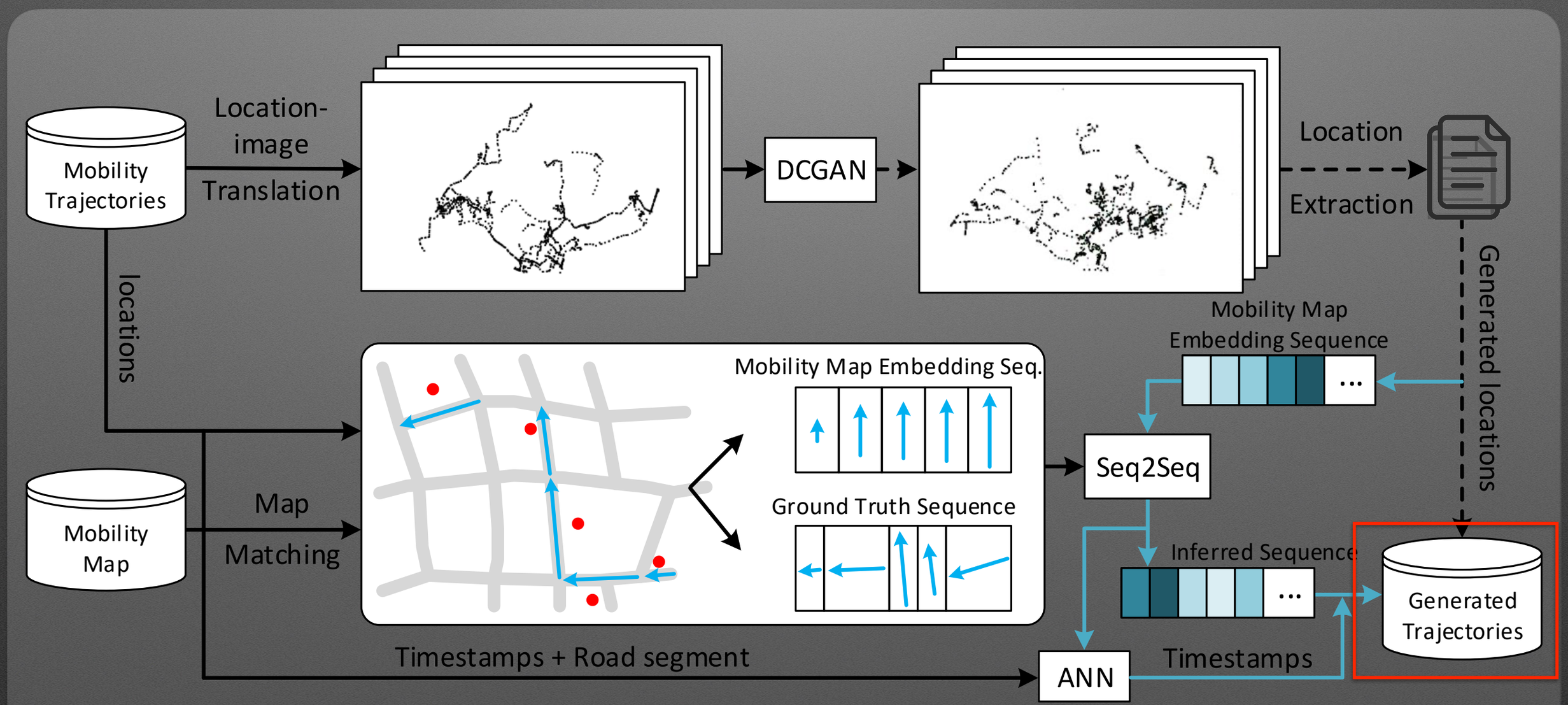
3. System Design

Generate new trajectories in TrajGen.



3. System Design

Generate new trajectories in TrajGen.



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

GANs are proved to perform well in image related tasks.

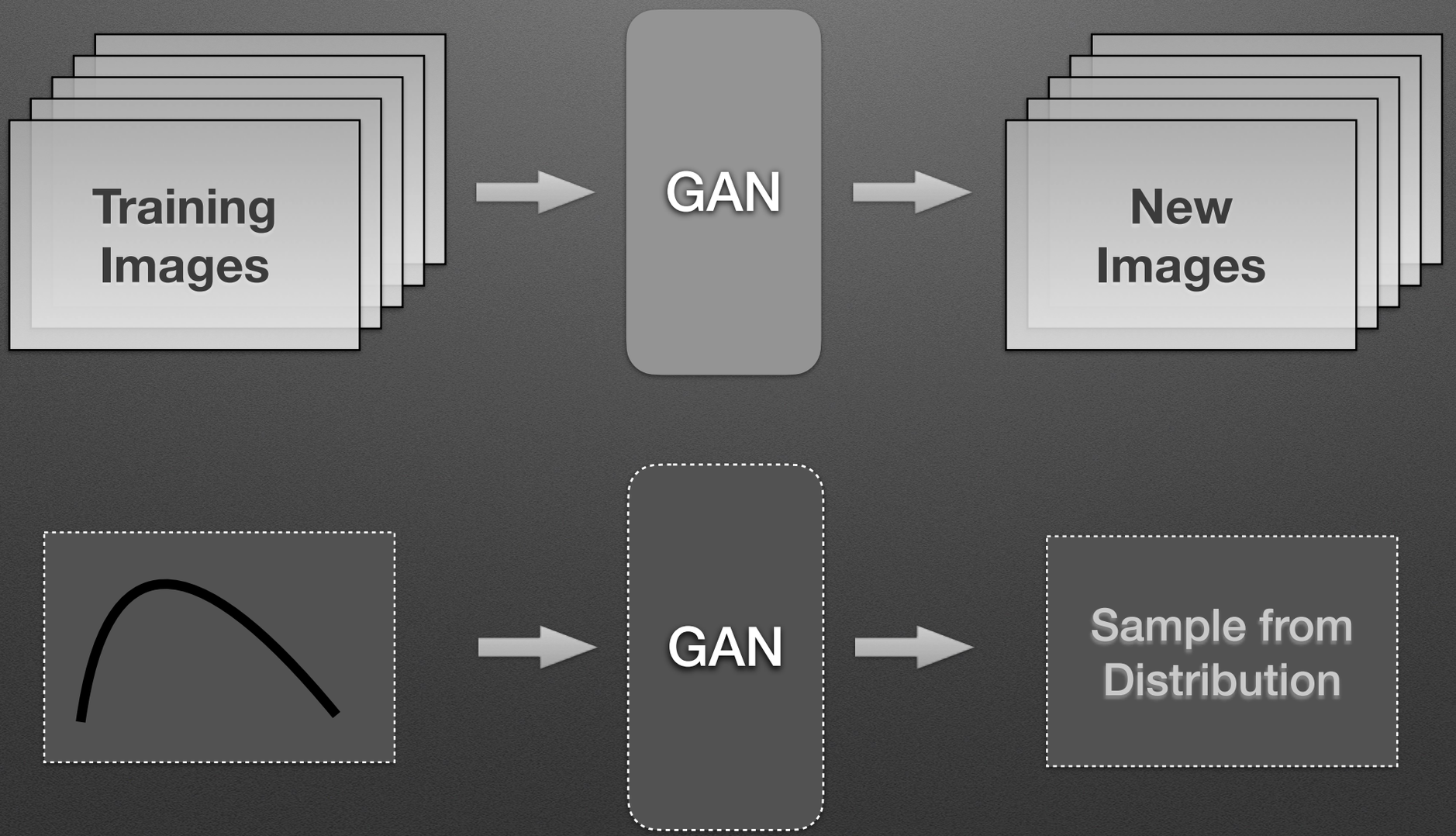
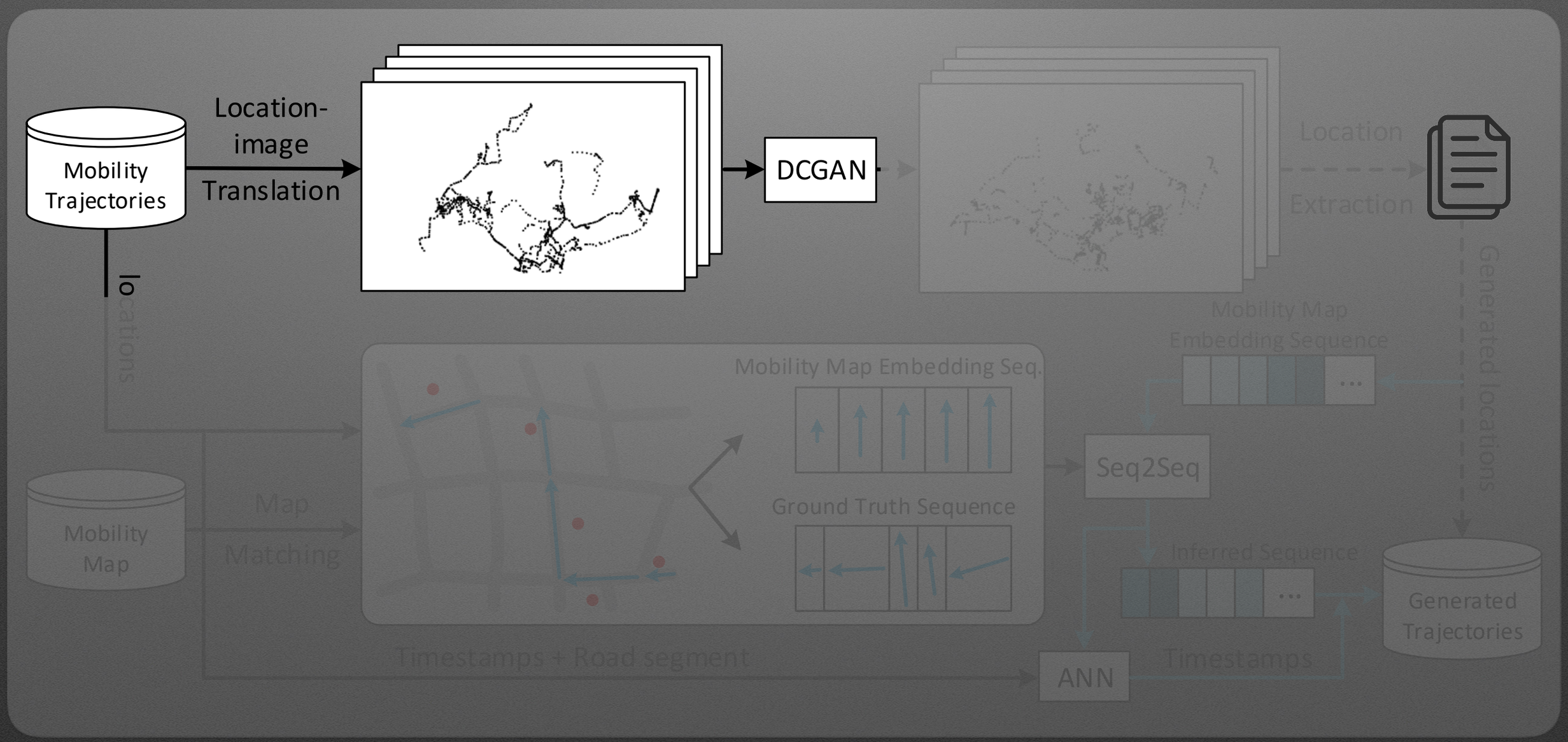


Image is a spatial distribution of pixels.

3. System Design

Spatial information learning

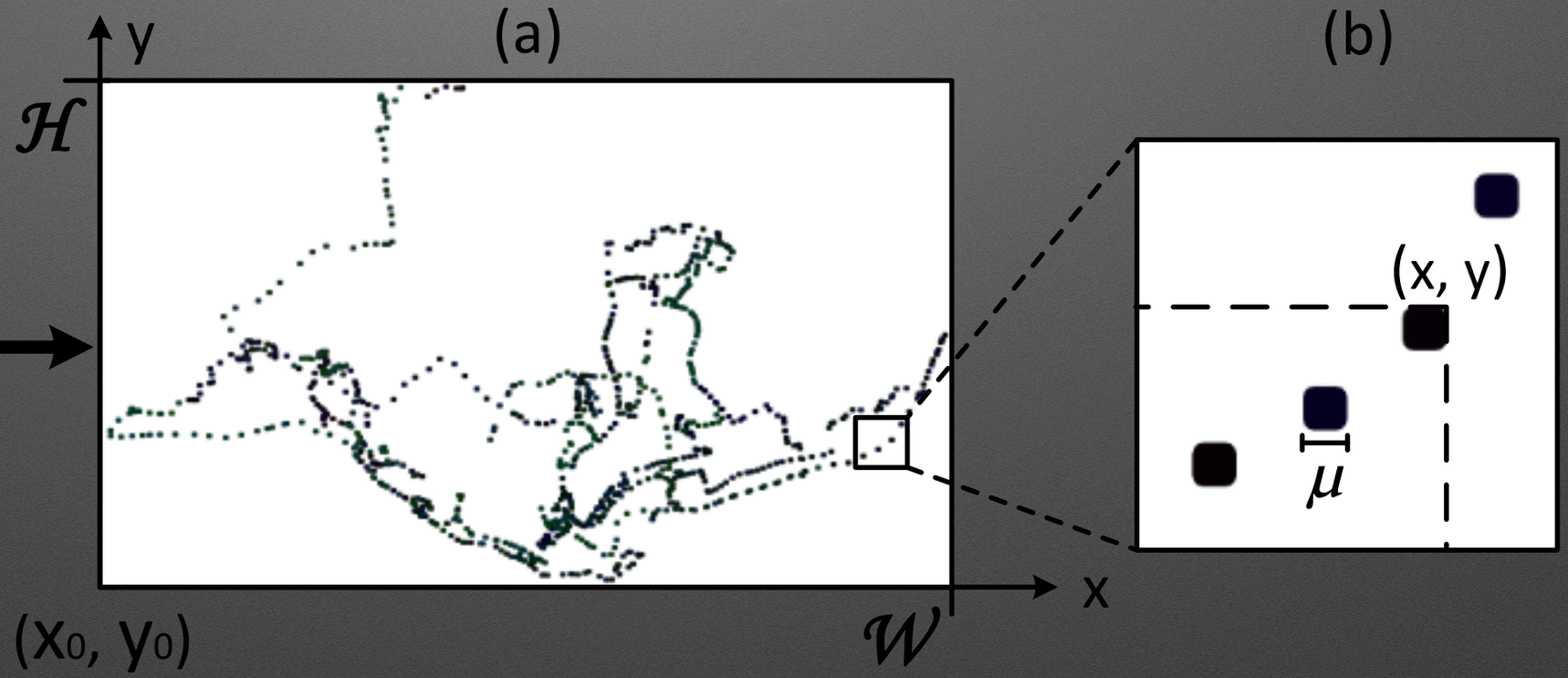


3. System Design

Spatial information learning | Location-image translation

Mobility Data

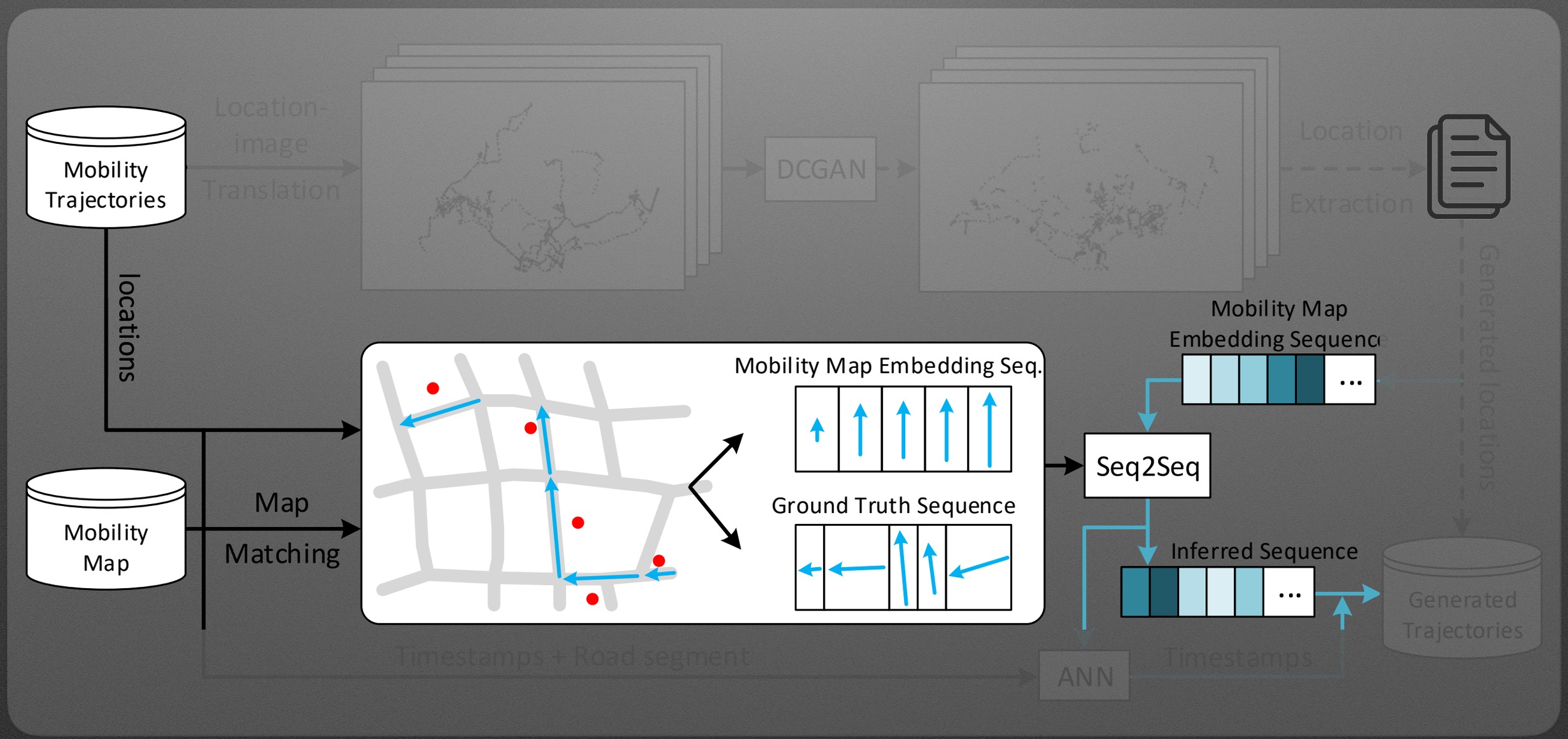
No.	Location
1	lat1, lon1
2	lat2, lon2
3	lat3, lon3
4	lat4, lon4
5	lat5, lon5
⋮	⋮



Ensure the distribution on image is the same as that in reality.

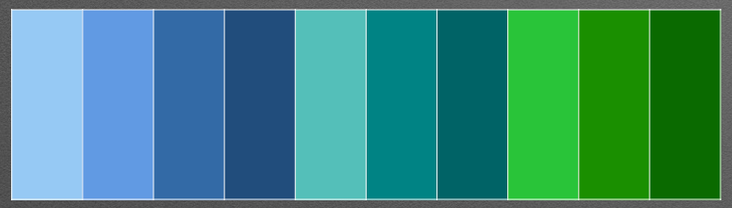
3. System Design

Sequence information learning



3. System Design

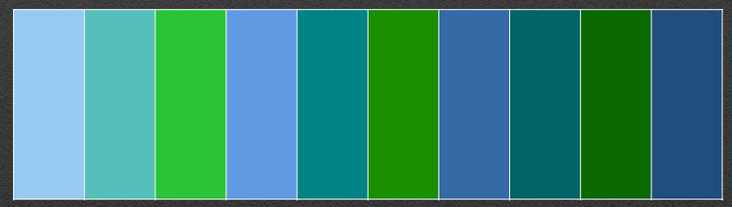
Temporal information can be regarded as a combination of sequence information and timestamps.



Road segment sequence

seq2seq

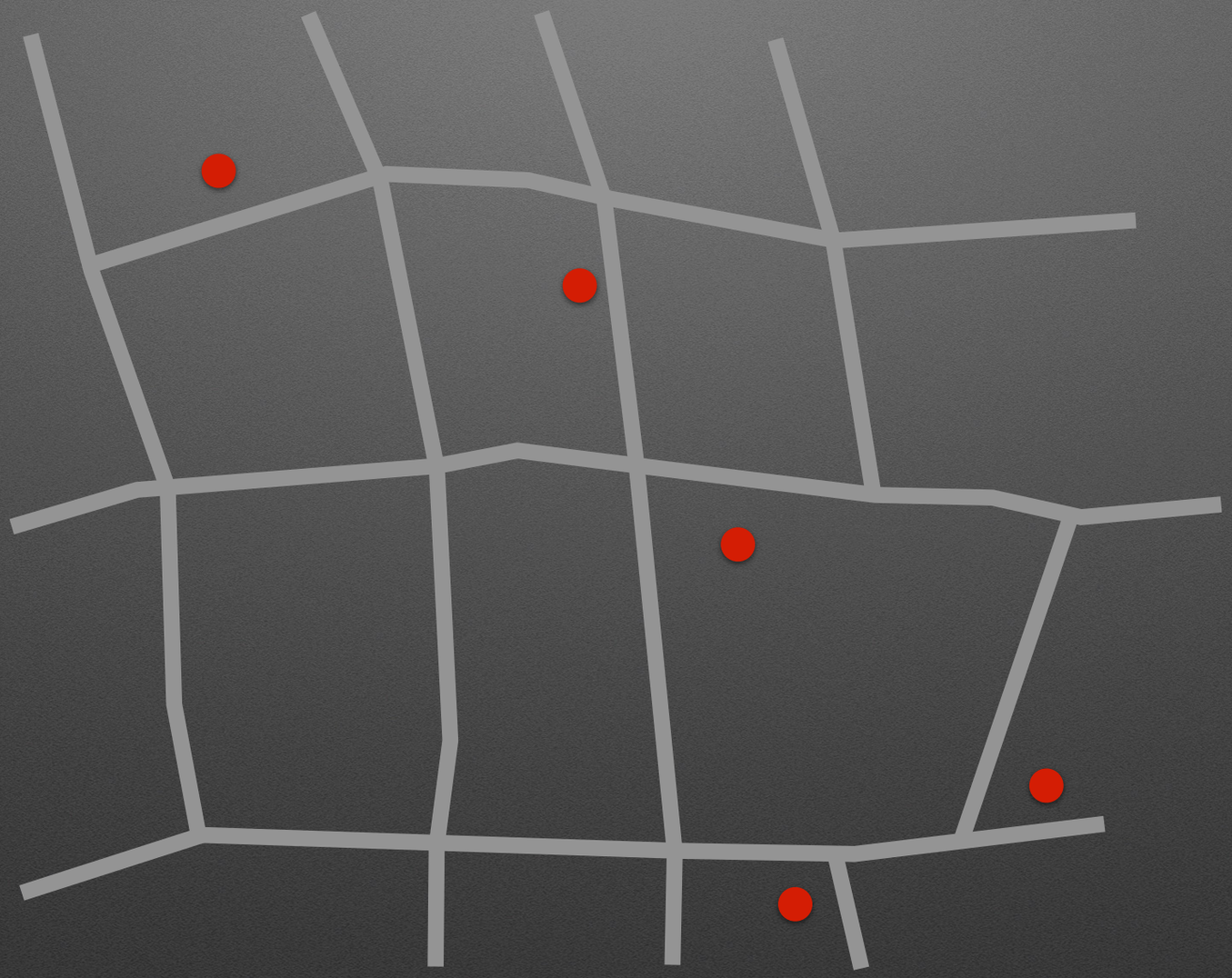
Travel patterns



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

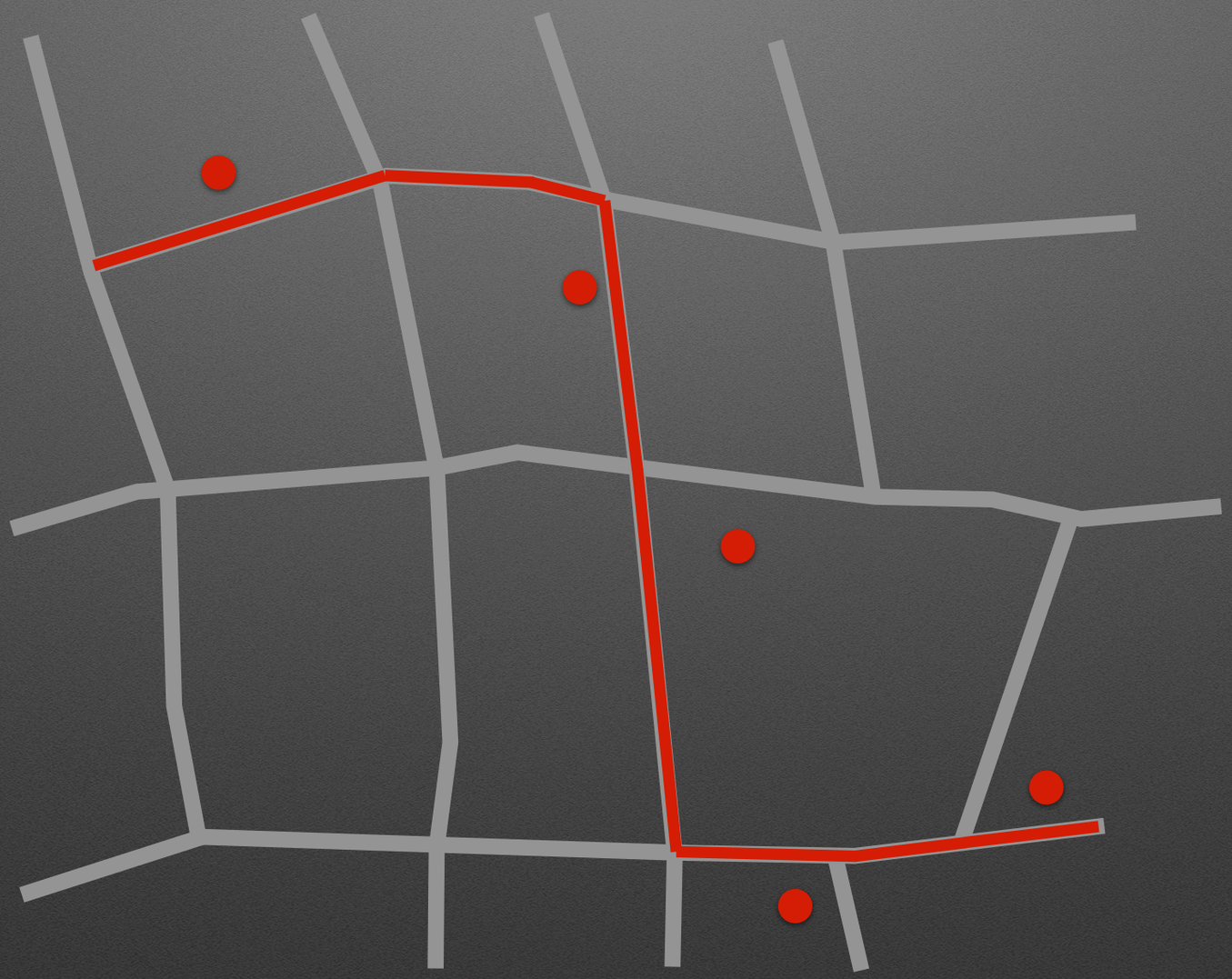
Sequence information learning | Mobility map embedding



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

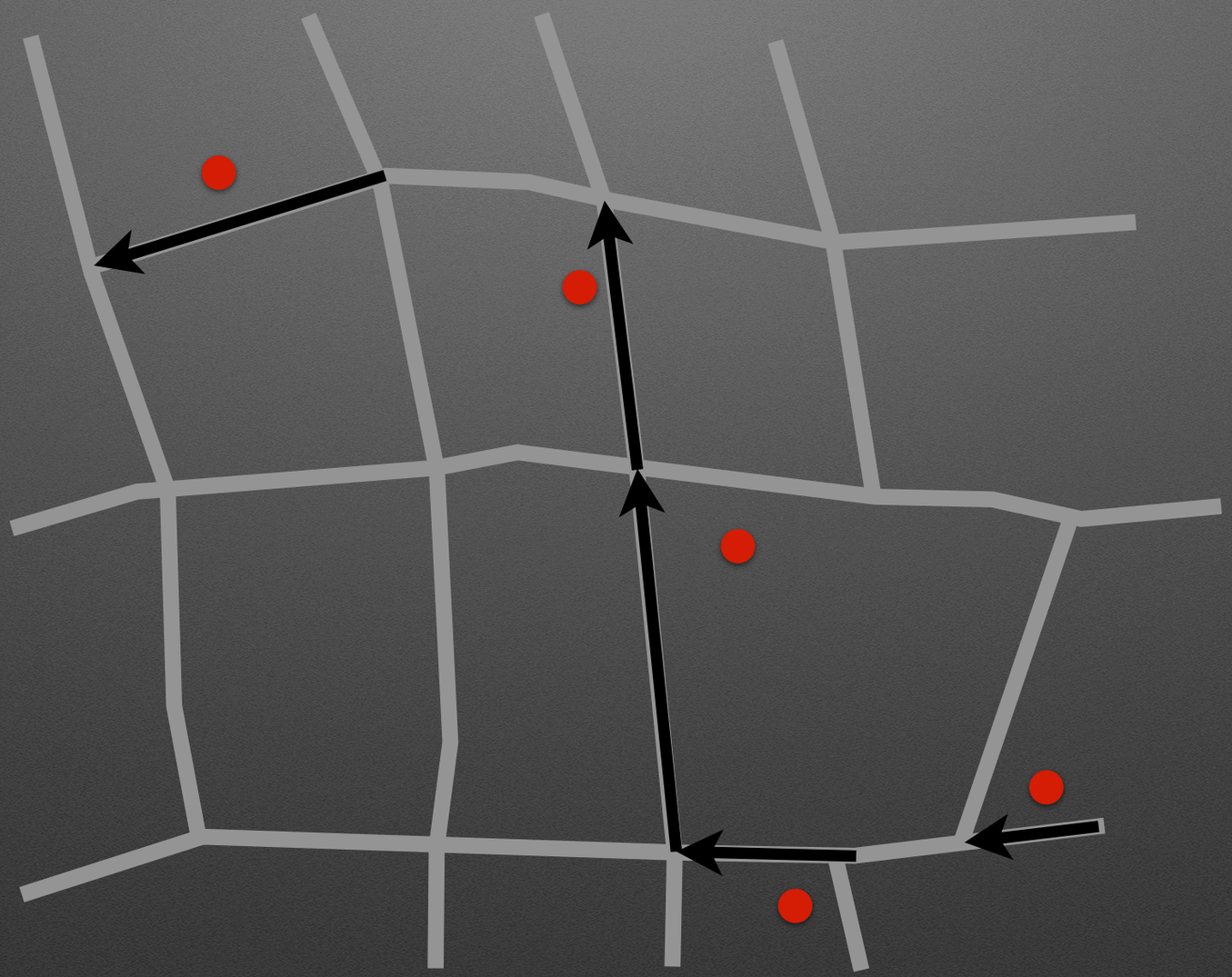
Sequence information learning | Mobility map embedding



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

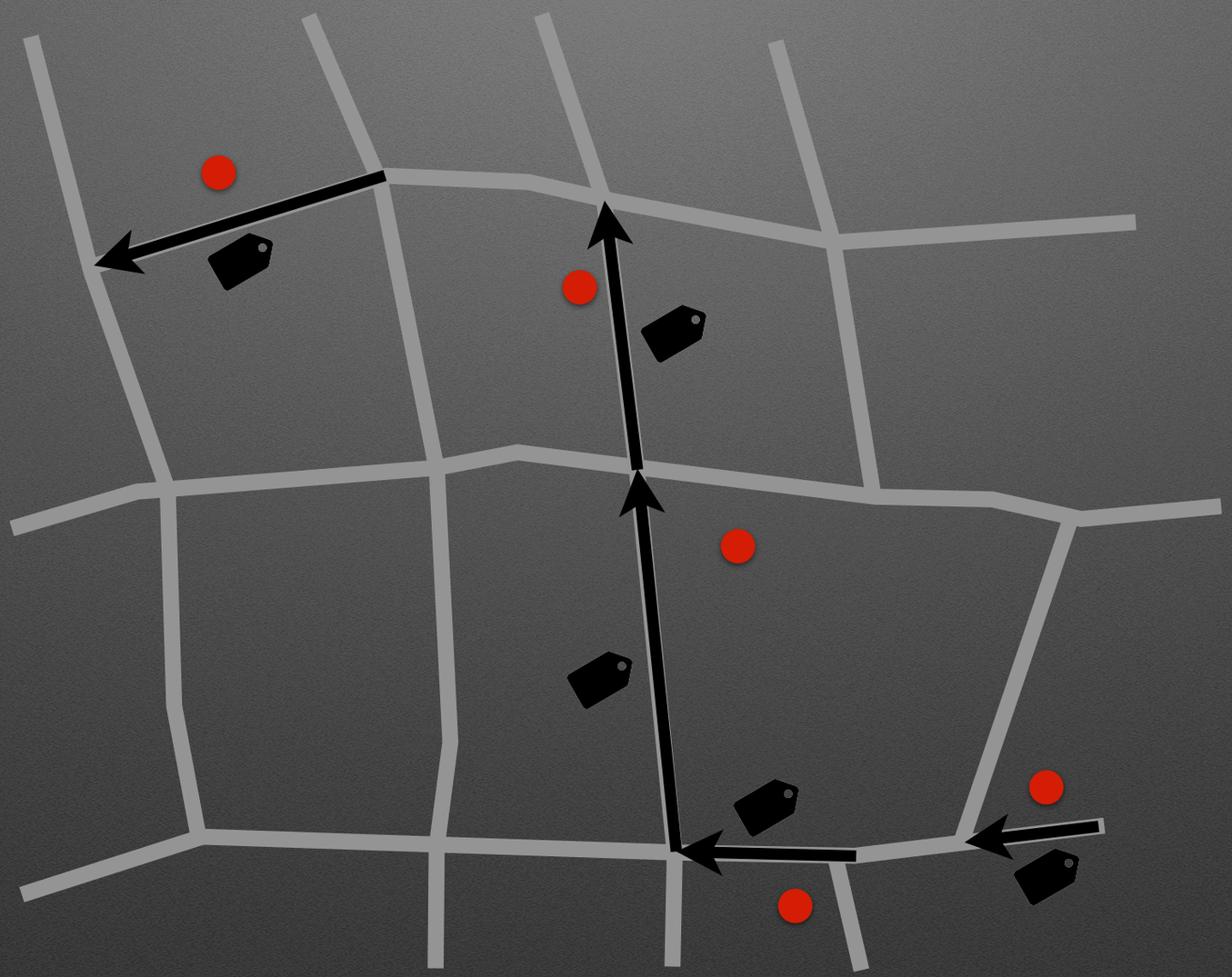
Sequence information learning | Mobility map embedding



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

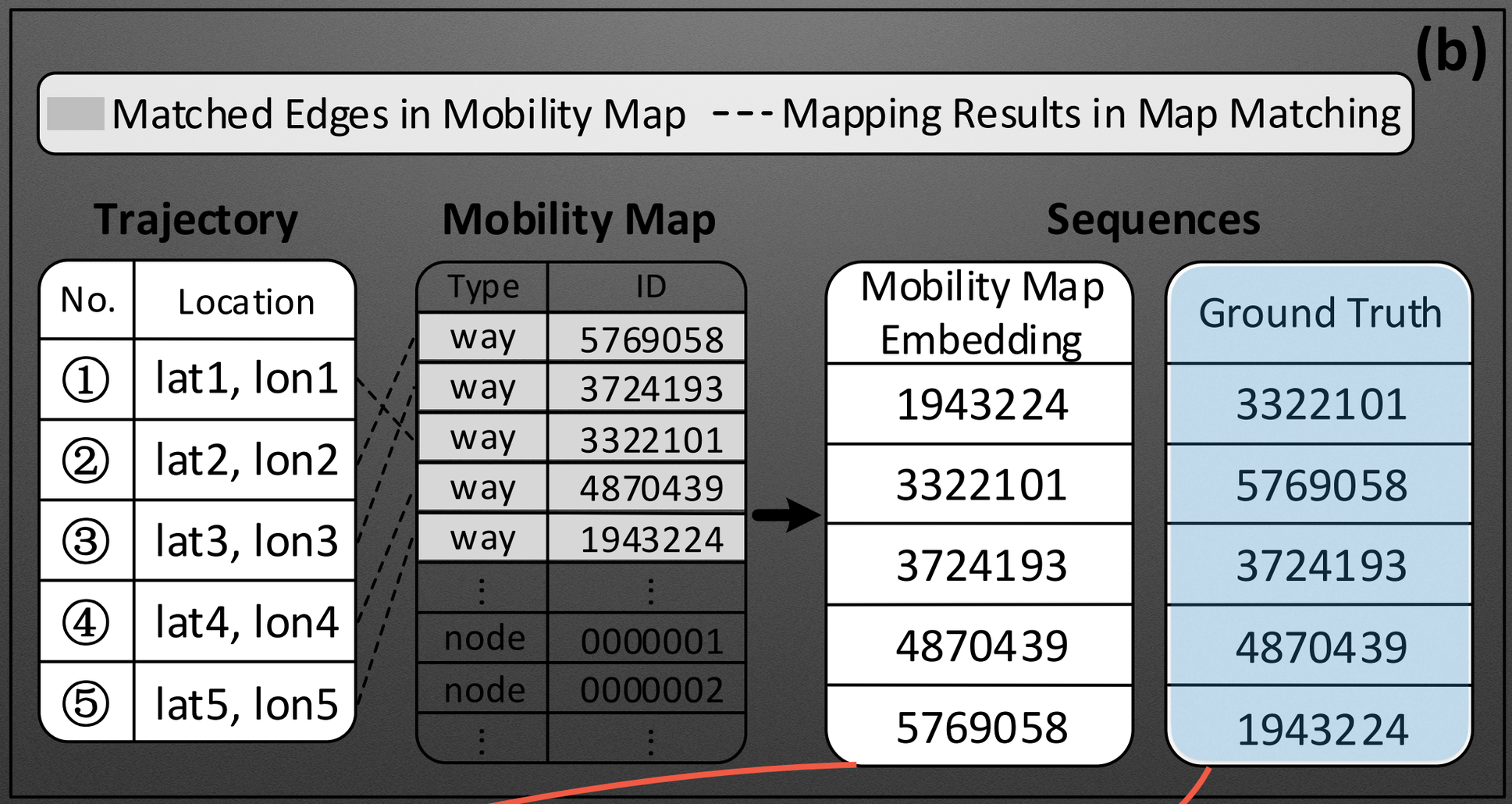
Sequence information learning | Mobility map embedding



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3. System Design

Sequence information learning | Mobility map embedding



Intrinsic Sequence + Ground Truth Sequence - -> Seq2Seq Model

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

4. Experiment

Dataset: Singapore Taxi Dataset

- # of taxis: 17,610 in total
- One sample per 30 second
- Last four months
- Cover the whole Singapore
- Hundreds of locations in each trajectory

Benchmarks:

Random Perturbation	Random shift Random direction
Gaussian Perturbation	Shift based on Gaussian Direction based on Gaussian
TrajGen-v	Generated by TrajGen

4. Experiment

Data preparation:

- Raw: randomly select 500 trajectories from original data
- RP: randomly select 500 trajectories from RP data
- GP: randomly select 500 trajectories from GP data
- TrajGen: randomly select 500 trajectories from TrajGen data
- TrajGen:-v randomly select 500 trajectories from TrajGen-v data

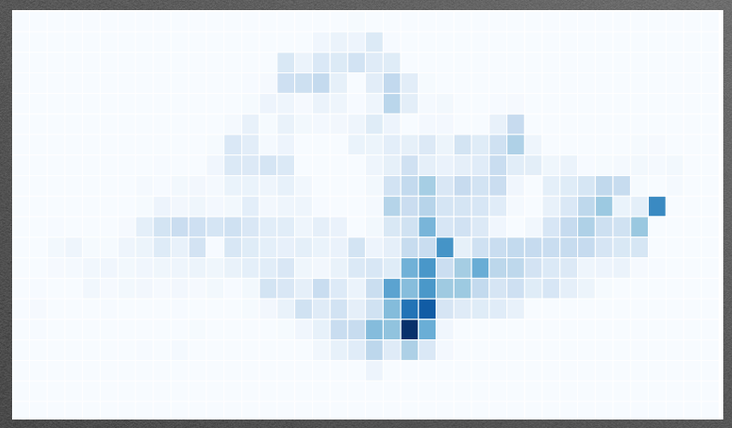
Considerations:

- Spatial distribution
- Temporal distribution
- Statistical features

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

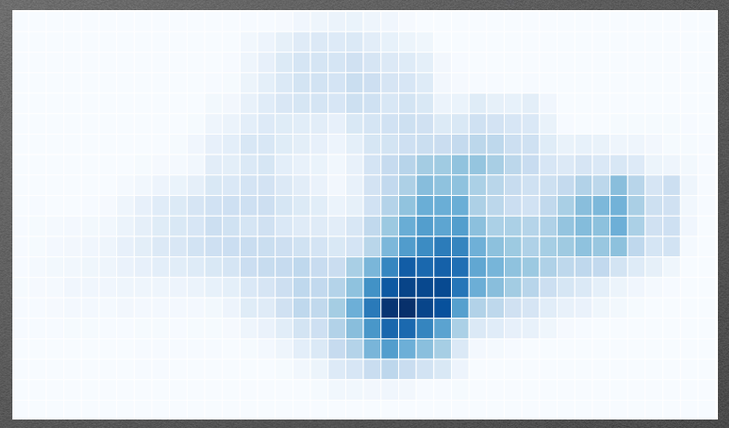
4. Experiment

Spatial distribution across Singapore.



Raw Data 1

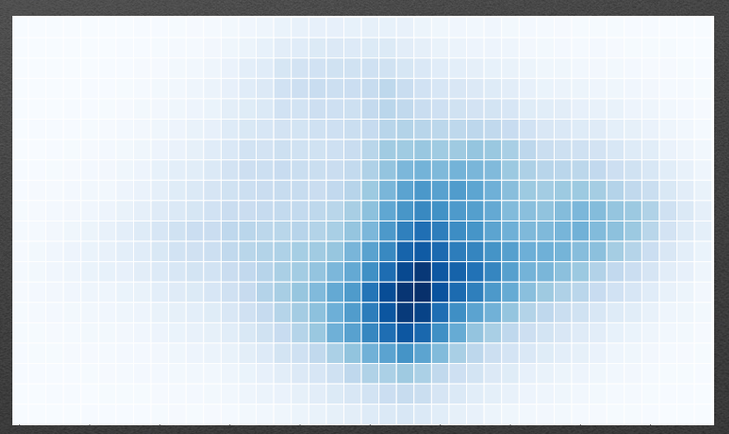
TrajGen Data



RP Data

Raw Data 2

TrajGen-v Data

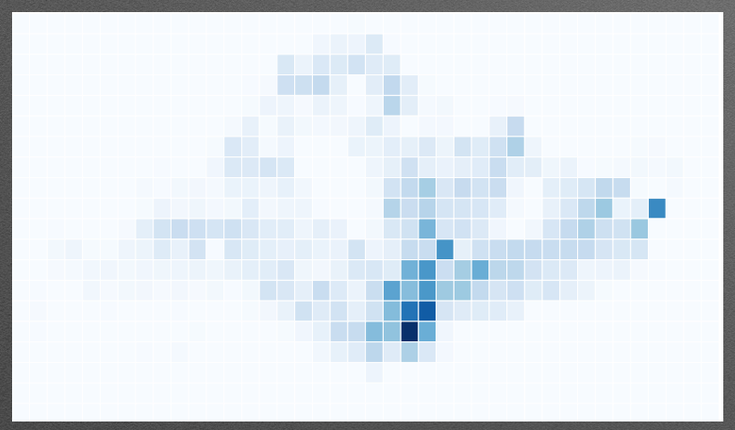


GP Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

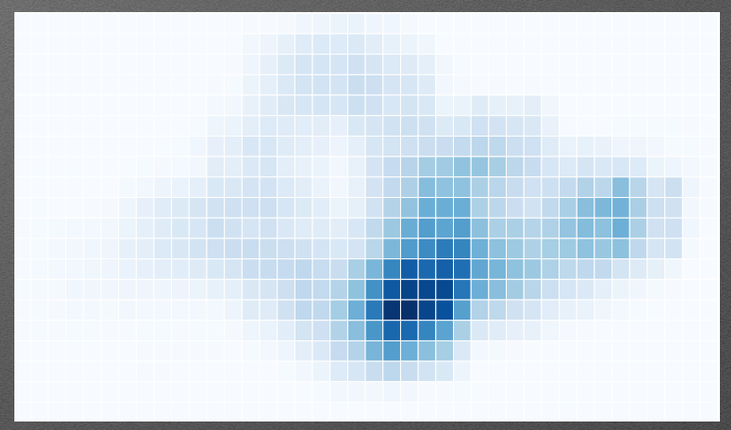
4. Experiment

Spatial distribution across Singapore.



Raw Data 1

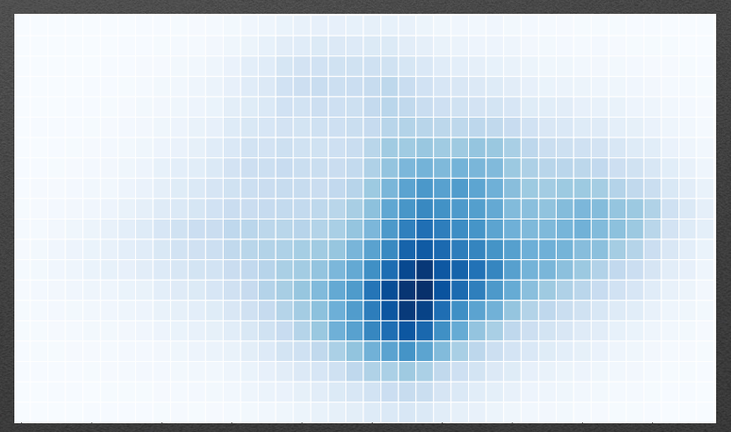
TrajGen Data



RP Data

Raw Data 2

TrajGen-v Data

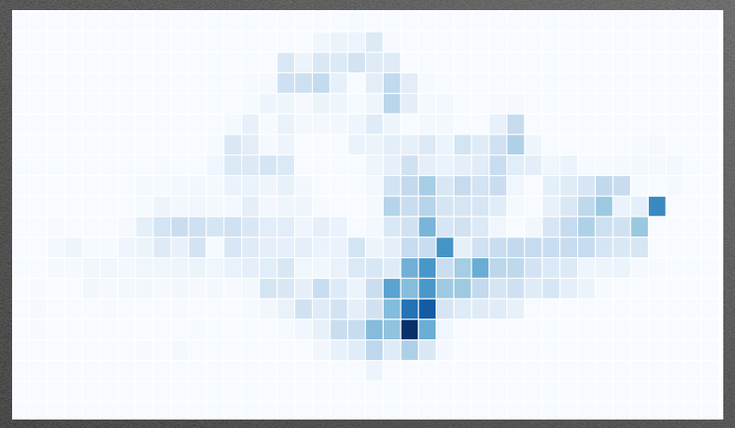


GP Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

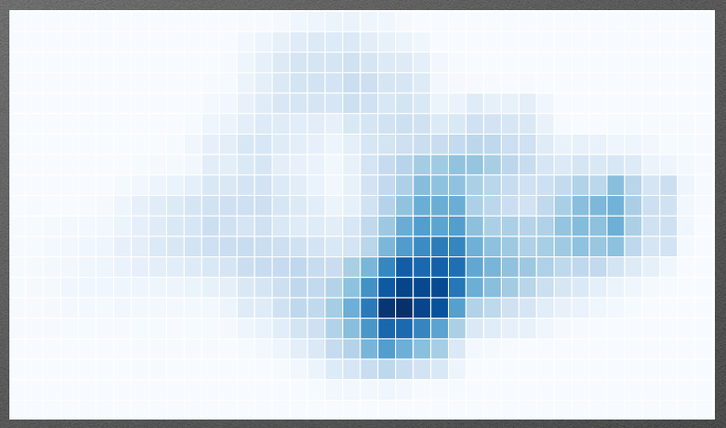
4. Experiment

Spatial distribution across Singapore.

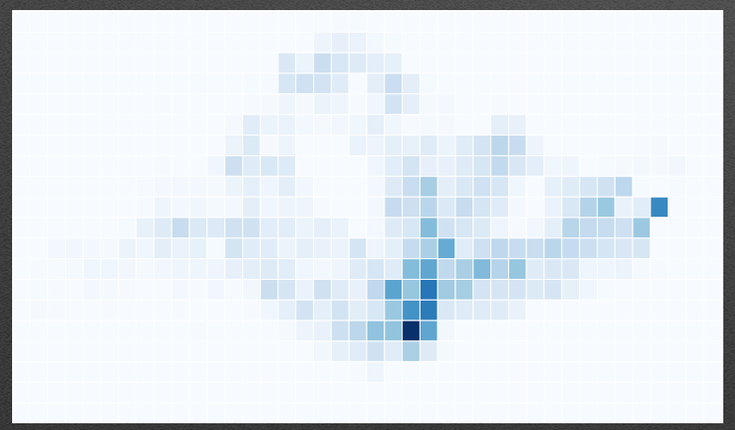


Raw Data 1

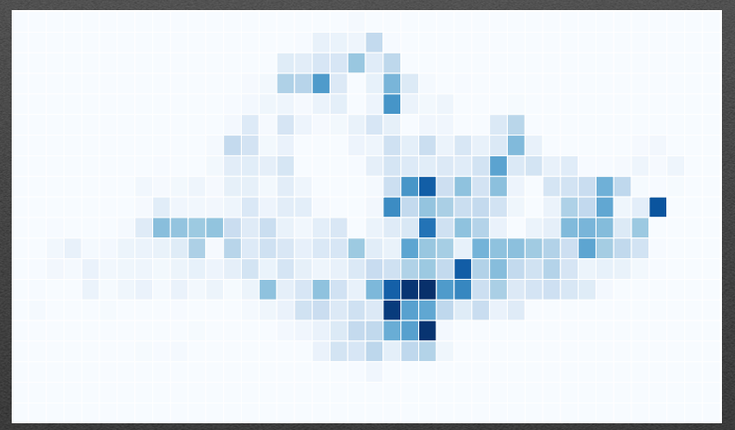
TrajGen Data



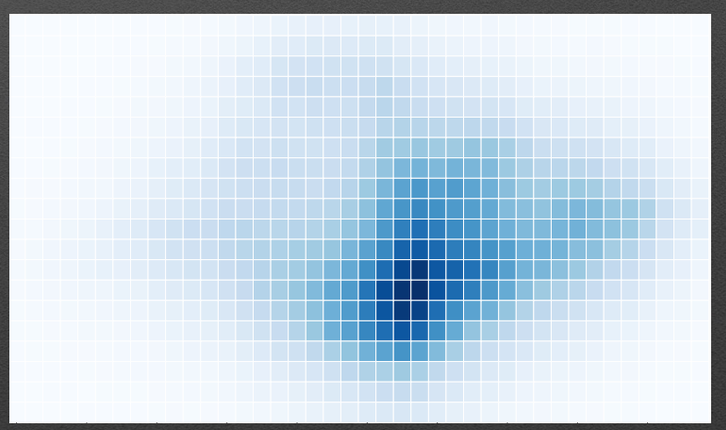
RP Data



Raw Data 2



TrajGen-v Data

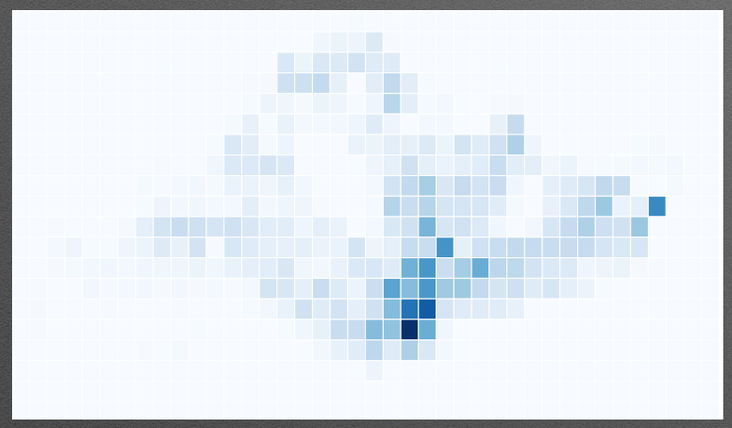


GP Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

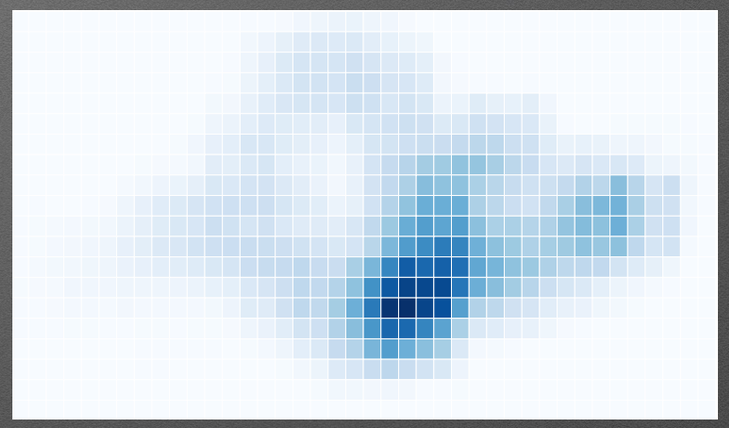
4. Experiment

Spatial distribution across Singapore.

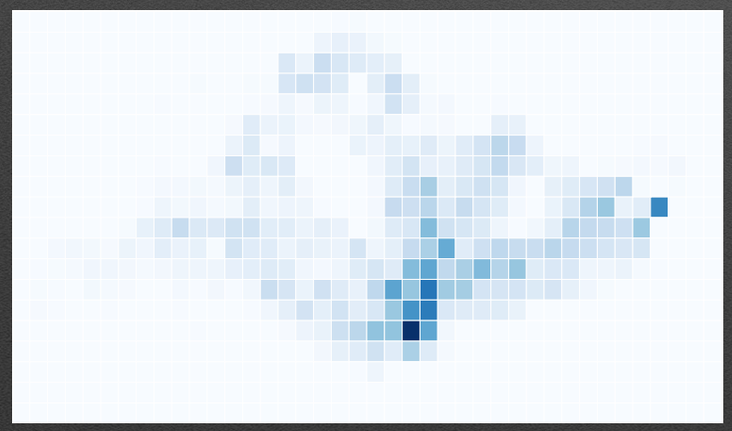


Raw Data 1

TrajGen Data

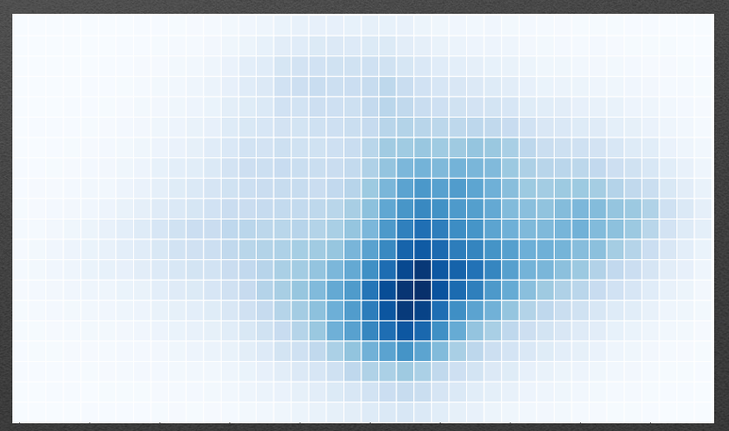


RP Data



Raw Data 2

TrajGen-v Data



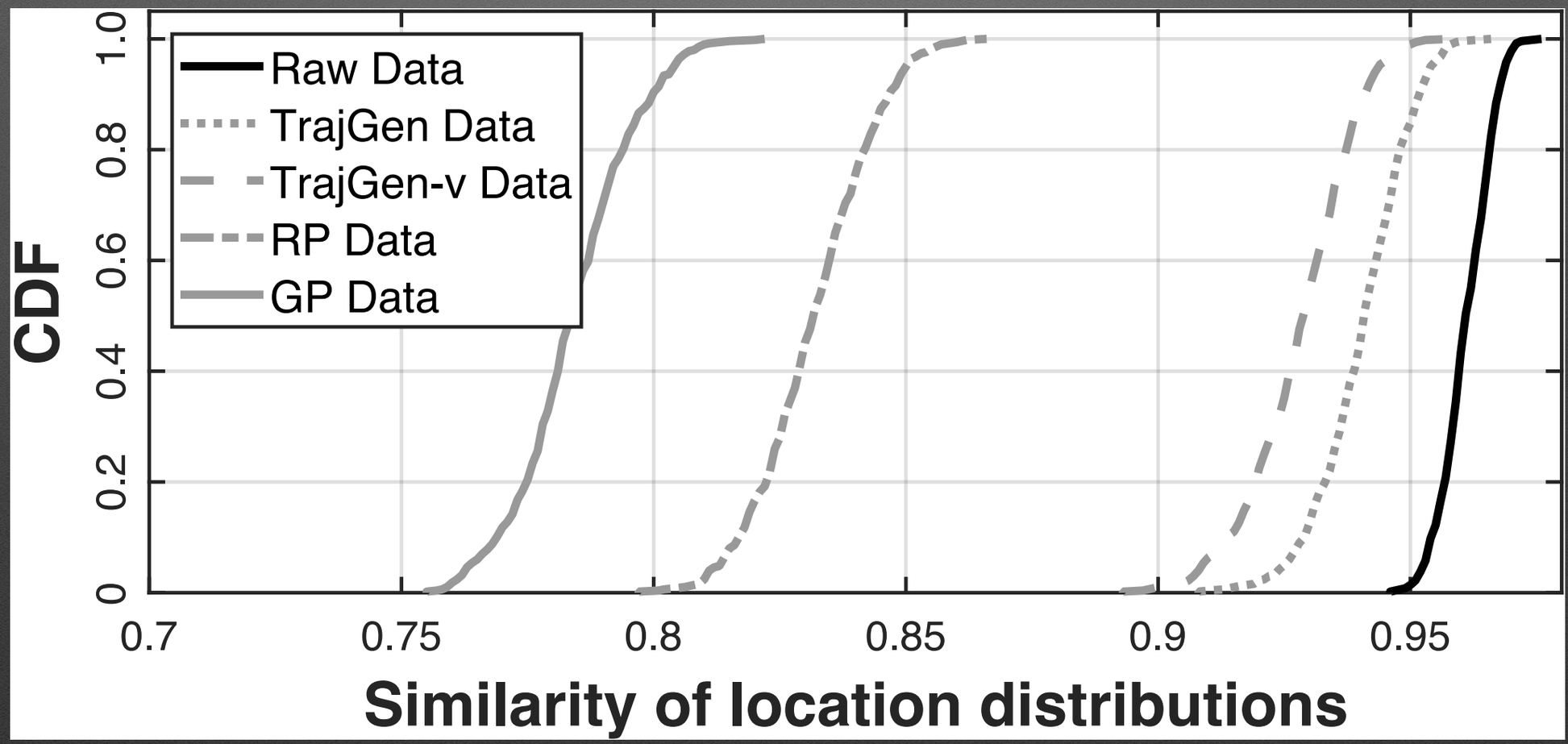
GP Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

4. Experiment

Spatial distribution across Singapore.

Repeat similarity computing 500 times.

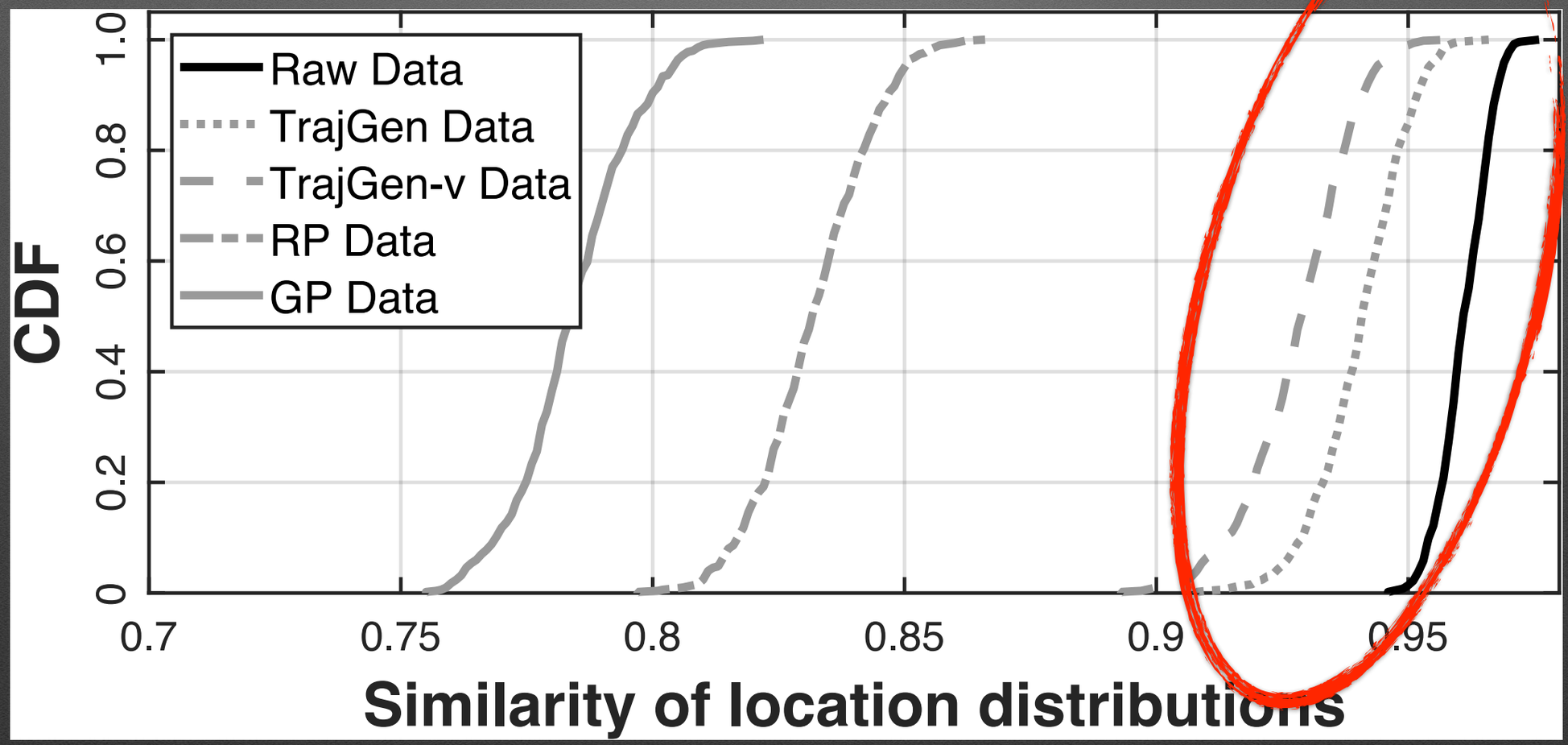


NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

4. Experiment

Spatial distribution across Singapore.

Repeat similarity computing 500 times.

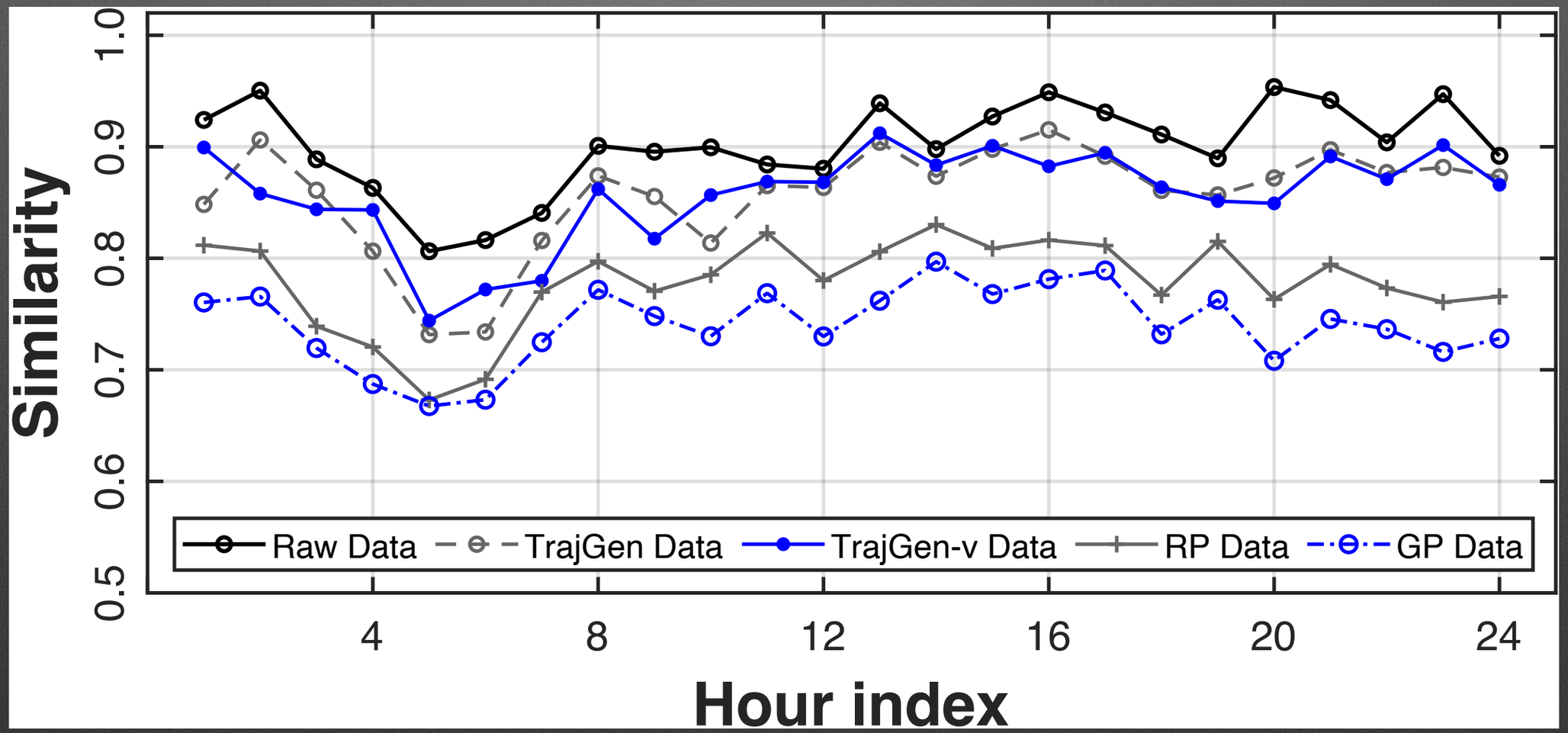


TrajGen could generate mobility trajectories that own similar spatial distribution with the original data.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

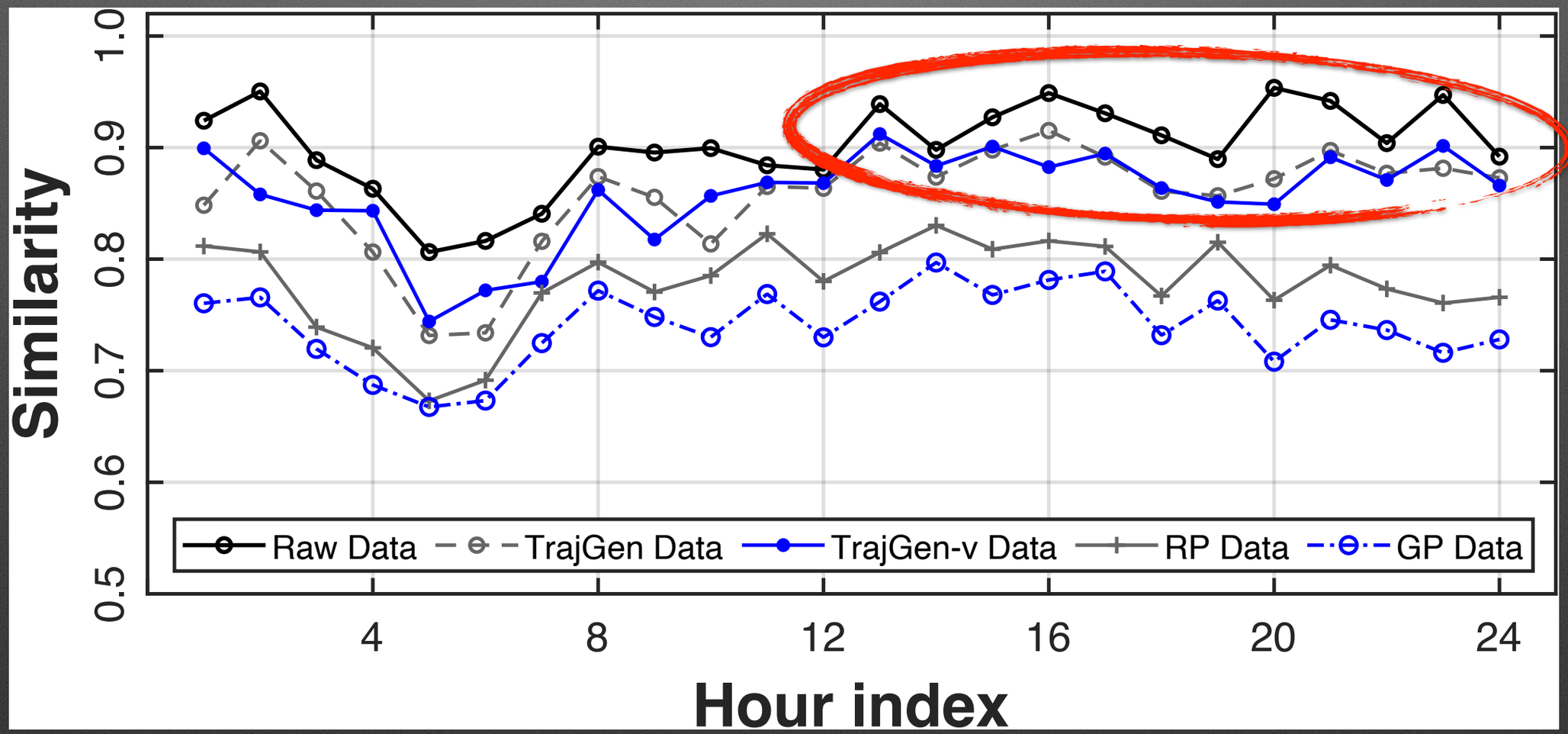
4. Experiment

Spatial distribution in different time slots.



4. Experiment

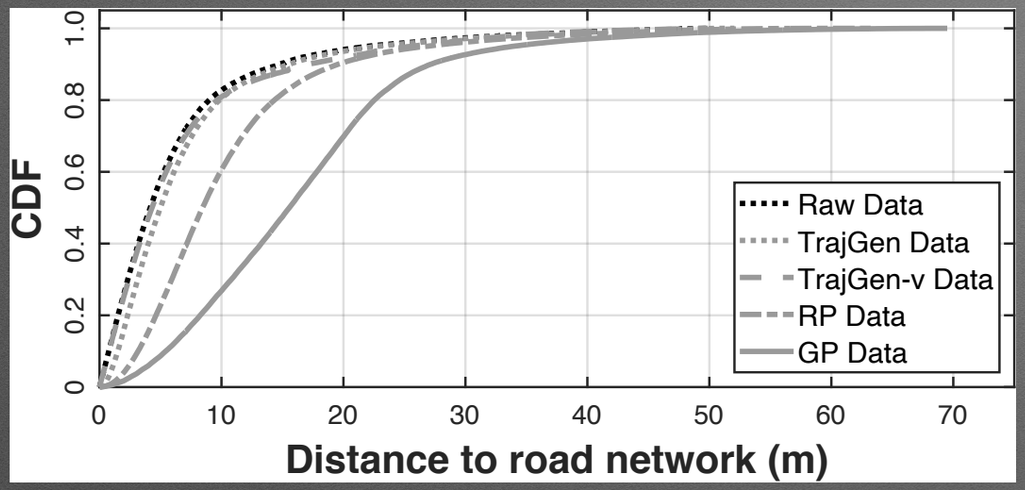
Spatial distribution in different time slots.



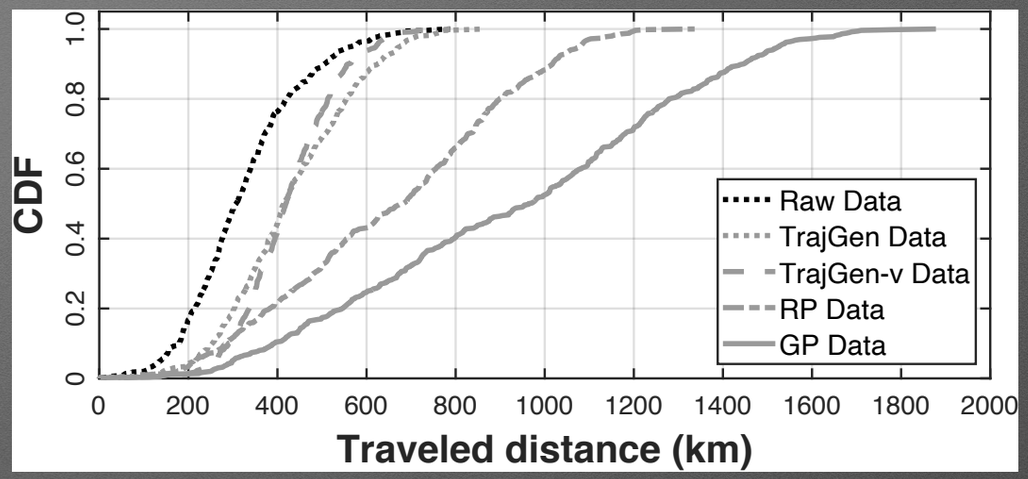
TrajGen could generate mobility trajectories that own similar spatial and temporal distribution with the original data.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

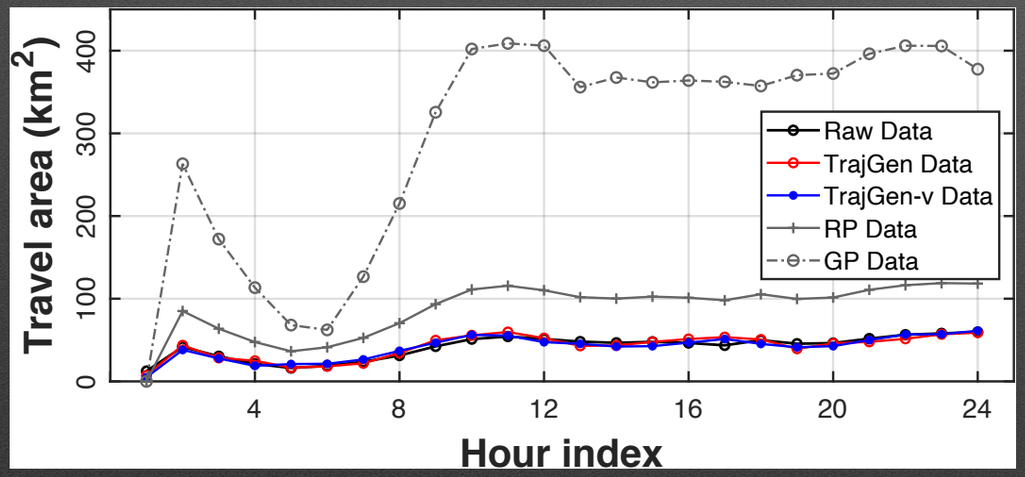
4. Experiment



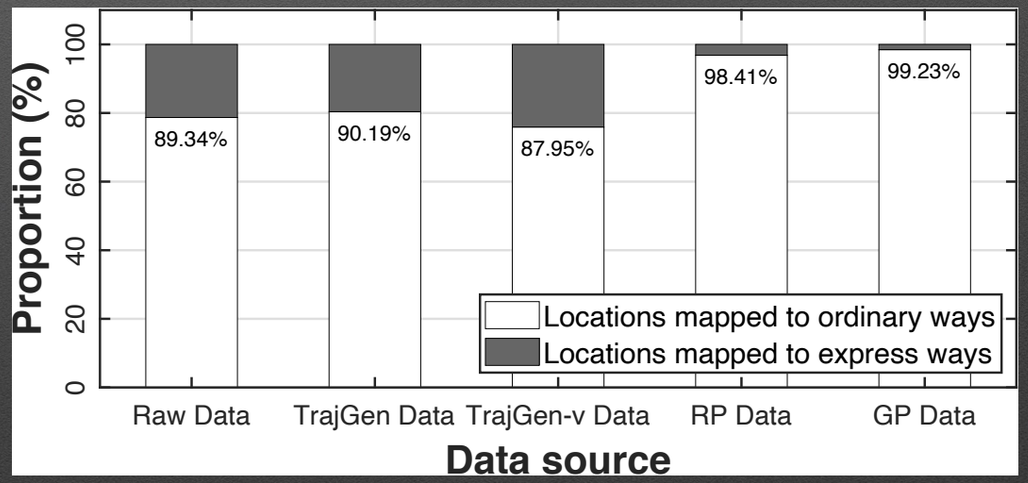
Distance from location to road



Travel distance of trajectories



Travel covered area



Ordinary way v.s. express way

TrajGen could generate mobility trajectories that have similar characteristics with the original data.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

5. Case Study

Road Discovery [1, 2, 3, 4]

[1] H. Wu, et al. GLUE: a Parameter-Tuning-Free Map Updating System. In ACM CIKM 2015.
[2] C. Cao, et al. Walkway Discovery from Large Scale Crwodsensing. In IEEE/ACM IPSN 2018.
[3] Z. Shan, et al. COBWEB: A Robust Map Updating System Using GPS Trajectories. In ACM UbiComp 2015.
[4] Y. Wang, et al. CrowdAtlas: Self-Updating Maps for Cloud and Personal Use. In ACM MobiSys 2013.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

5. Case Study

Road Discovery

Google Maps

5. Case Study

Road Discovery

Raw Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

5. Case Study

Road Discovery

TrajGen Data

RP Data

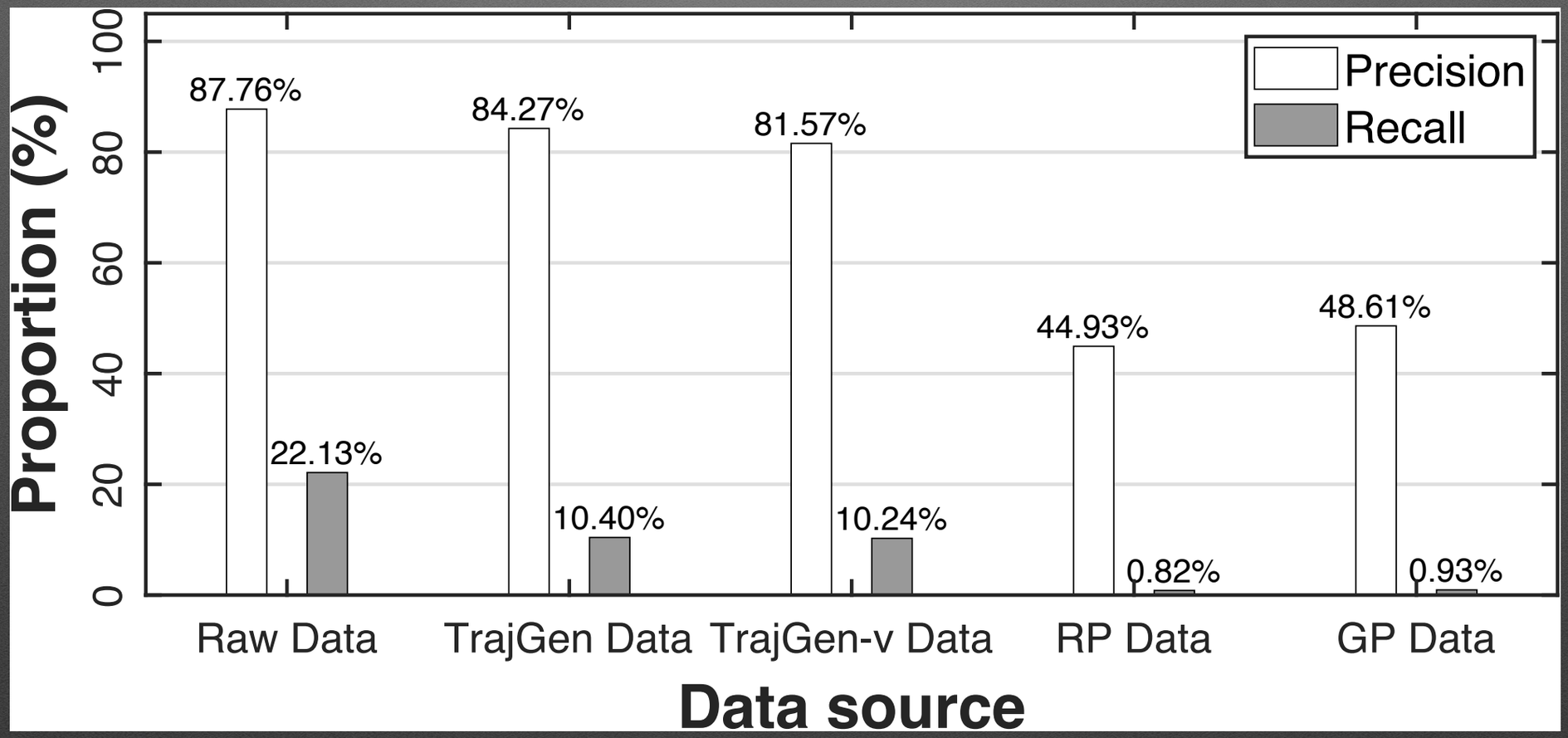
TrajGen-v Data

GP Data

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

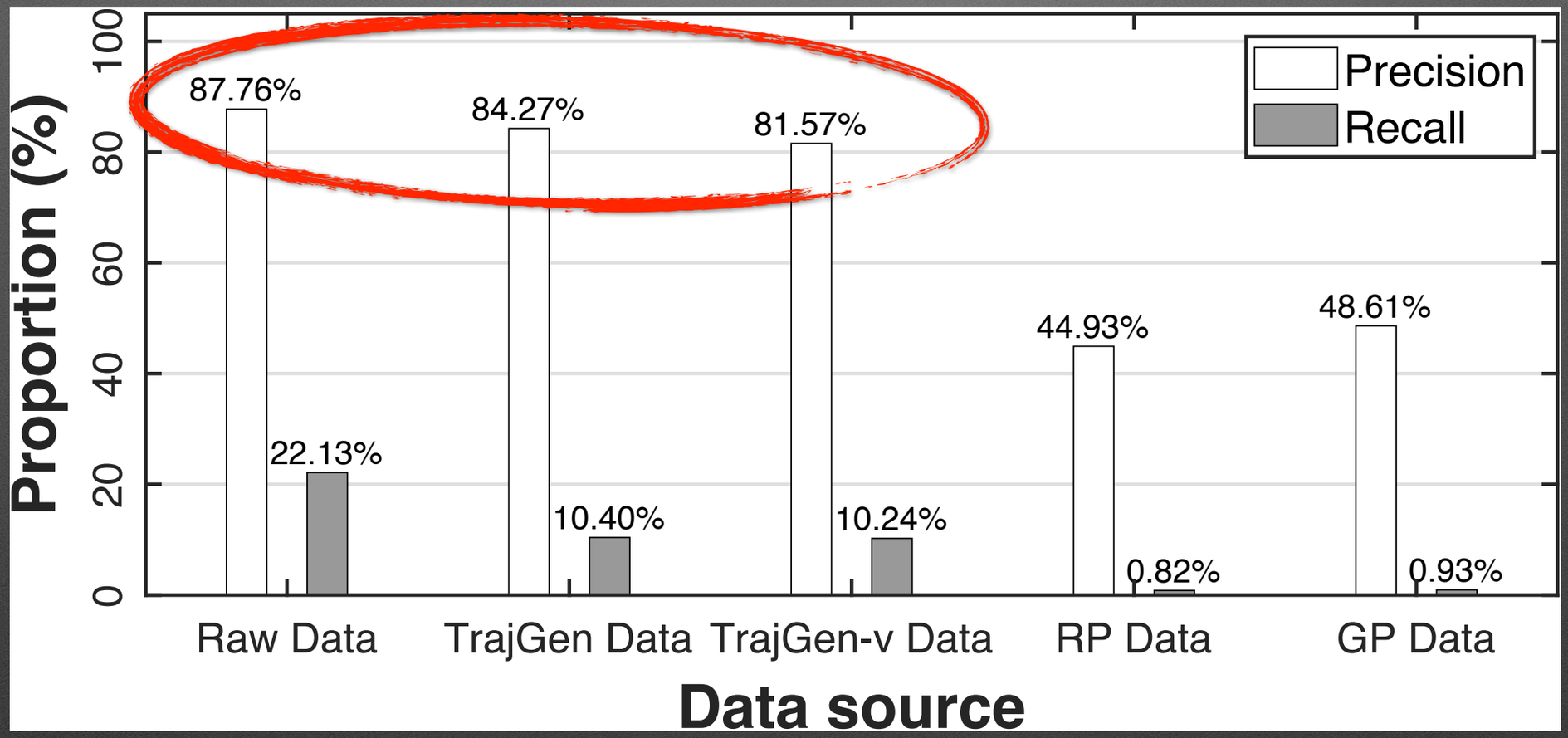
5. Case Study

Road Discovery



5. Case Study

Road Discovery

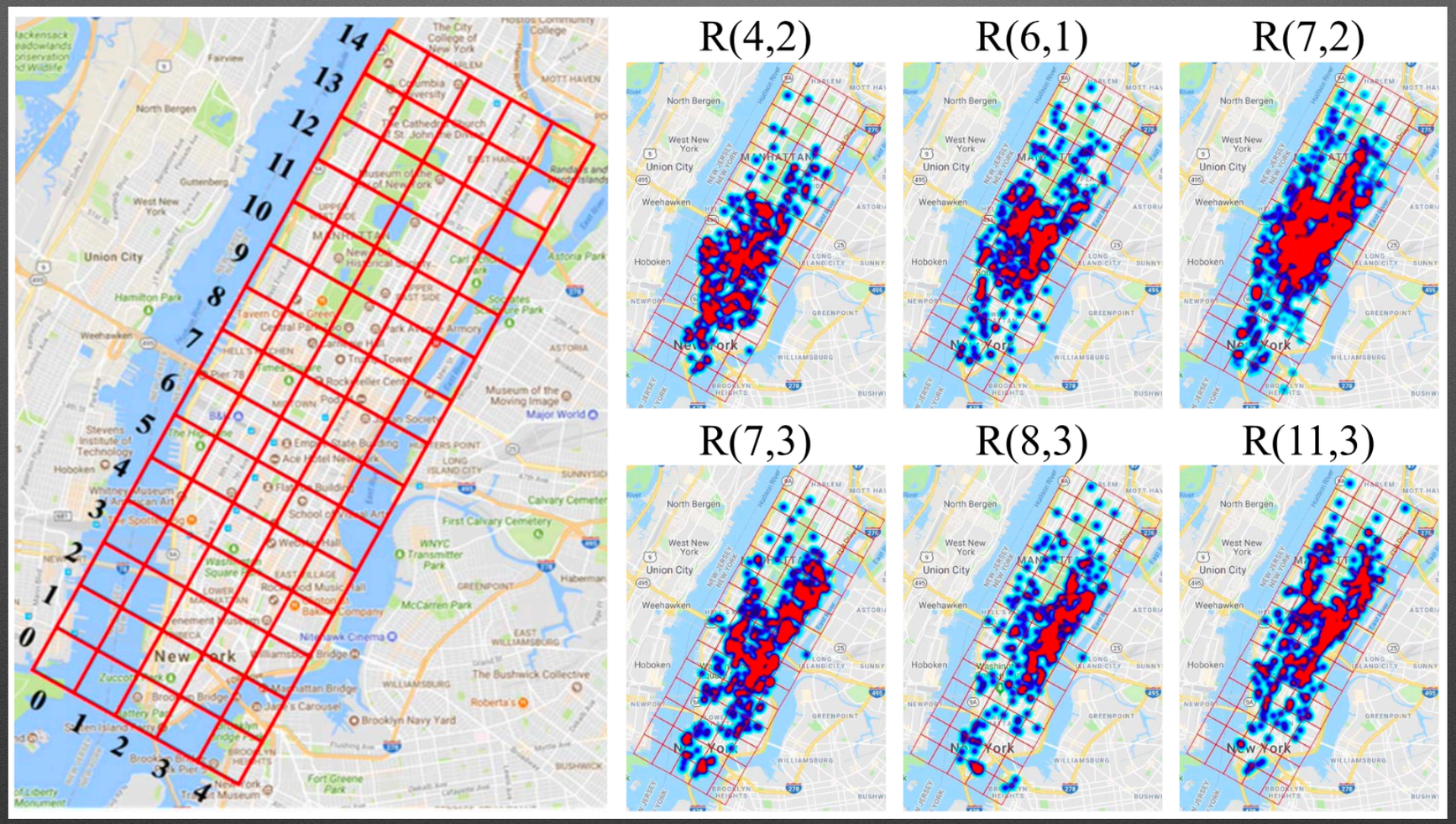


TrajGen could generate mobility trajectories that have similar performance with the original data in Road Map Updating application.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

5. Case Study

Origin-Destination Estimation [1, 2, 3, 4].

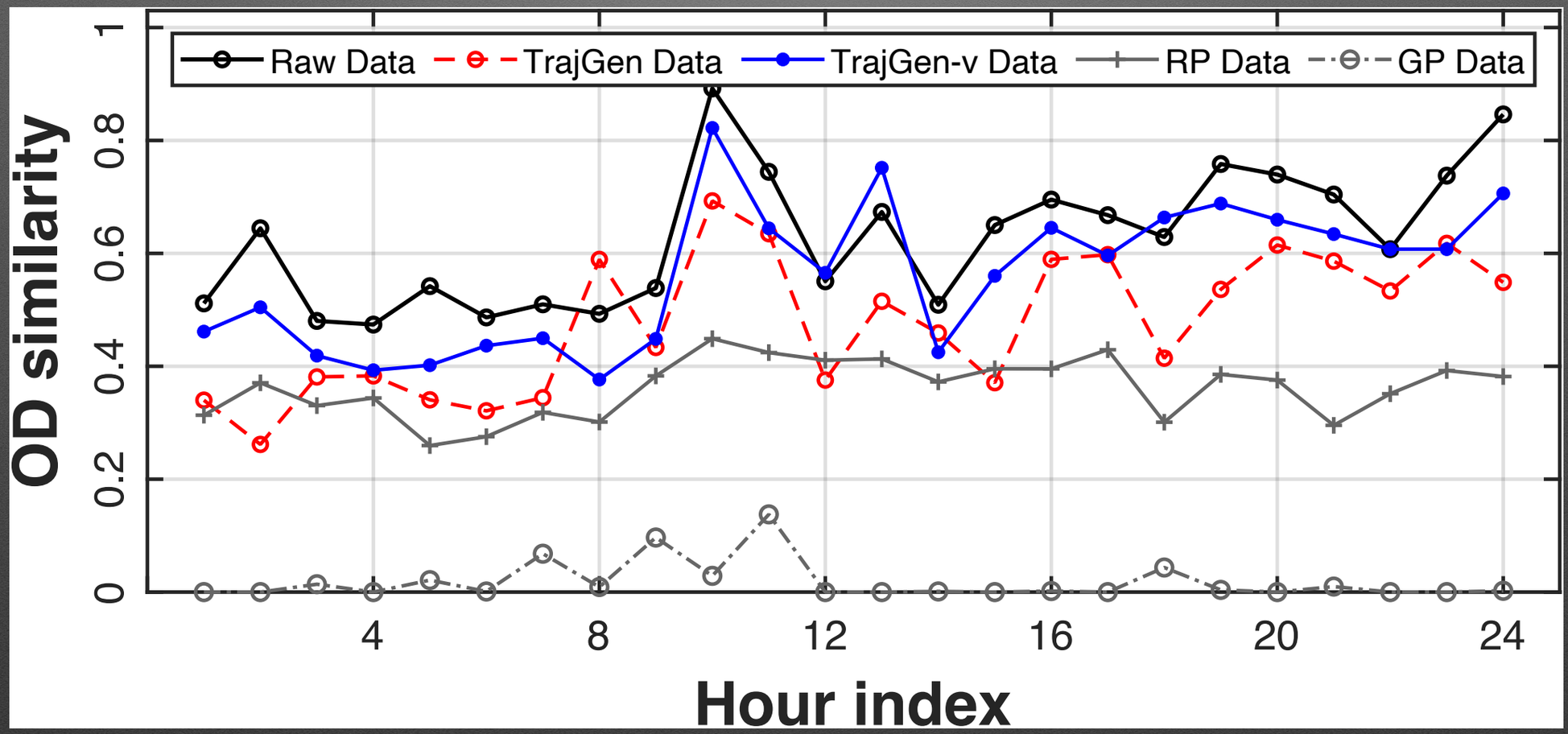


[1] L. Liu, et al. Contextualized Spatial-Temporal Network for Taxi Origin-Destination Demand Prediction. In IEEE TITS 2019.
[2] K. Zhang, et al. A Framework for Passengers Demand Prediction and Recommendation. In IEEE CSC 2016.
[3] J. Xu, et al. Real-Time Prediction for Taxi Demand Using Recurrent Neural Networks. In IEEE TITS 2018.
[4] A. Anwer, et al. ChangiNow: A Mobile Application for Efficient Taxi Allocation at Airports. In IEEE ITSC 2013.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

5. Case Study

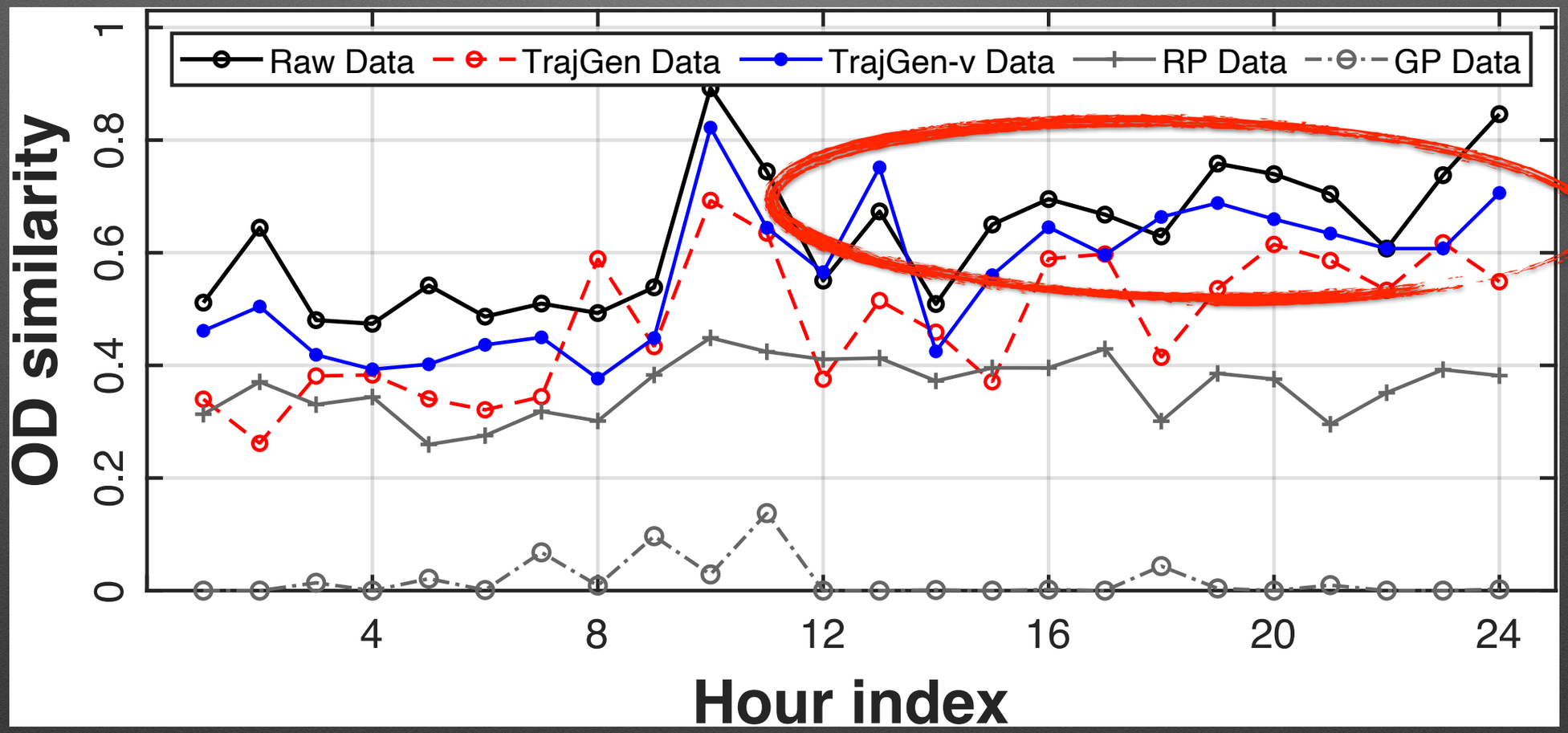
Origin-Destination Estimation.



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

5. Case Study

Origin-Destination Estimation.



TrajGen is able to generate artificial mobility trajectories owing similar OD demand with the original data across different hours.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Q & A

Thank you very much.