

Generating Human Mobility Route Based on Generative Adversarial Network

Ha Yoon Song
Department of Computer Engineering
Hongik University
Seoul, Republic of Korea
Email: hayoon@hongik.ac.kr

Moo Sang Baek
Research Institute of
Science and Technology
Hongik University
Seoul, Republic of Korea
Email: moosangbaek@gmail.com

Minsuk Sung
Department of Computer Engineering
Hongik University
Seoul, Republic of Korea
Email: mssung94@mail.hongik.ac.kr

Abstract—Recently, many researches on human mobility are aiming to suggest the personal customized solution in the diverse field, usually by academia and industry. Combined with deep learning methods, it is able to predict and generate novel routes of objects from the mobility data including the given past trends. In this work, Generative Adversarial Network (GAN) model is introduced for creating individual mobility routes based on sets of accumulated personal mobility data. The mobility data had been collected by use of geopositioning system and personal mobile devices. GAN has Discriminator and Generator which are composed of neural networks, and can train and extract geopositioning information. A sequence of longitude and latitude can be geographically mapped, and matrices including all these information can be handled by GAN. The GAN-based model successfully handled individual mobility routes in this way. Consequently, our model can generate and suggest unexplored routes from the existing sets of personal geolocation data.

I. INTRODUCTION

INDIVIDUAL mobility data has a huge capacity and can provide significant information and knowledge to modern industries. By processing the result of analyzed mobile dataset, enterprises can get human personalities so that they can interact with their customers effectively and set directions in the field of marketing. For instance, the analysis of the consumers' data can help to determine the location of new commercial shop branches and to find the intersected location where their customers visit in common. Therefore, the companies can provide service improving customer satisfaction through the individual preference which is extracted from correlation among distinguished locations analyzed from mobility data.

Normally, personal location data can be divided into location clusters according to the data distribution, and movement patterns in each cluster vary depending on the individual's purpose and desire. It means that the location clusters imply individual lifestyles. By analyzing the correlation between the major location clusters, we could deduce the past visited points and predict the future visitations. This correlation is an important factor which is necessary for creating a new mobility route. Indeed, an appropriate data pre-processing is

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2019R1F1A1056123).

required for training GAN with human location data. In data pre-processing work, we have partitioned personal mobility data, and utilized 4 layers of Convolutional Neural Networks (CNNs) to deliver features of the partitioned data. In data post-processing work, we adopted CNN to preserve essential location feature in the form of latitude and longitude by improving GAN performance. We introduce GAN for generating mobility routes. In this GAN, there are two representative Discriminator and Generator networks. These networks compare geopositioning features extracted from each mobility data with the generated probability distributions, while backpropagating the differences in each network. Discriminator learns the features extracted from the whole individual mobility route with extra mobility route, and improves ability to discriminate the fake route obtained from Generator and the real route from input dataset. Consequently, the self-creating GAN network, which ensures to enhance the discriminative features, can generate unexplored routes for area size of 3 square kilometer.

Section II will discuss the related researches on human mobility data with diverse approaches. In section III, we analysis our raw data, and section IV explains method we use on the data pre-processing and post-processing. Section V describes detailed methodology with proposed GAN. In section VI presents the result of generation route with our model. Section VII concludes our research and discusses about our future research.

II. RELATED WORKS

In this section, we review previous investigations on human mobility patterns, which were focused on creating novel mobility route or next location by data mining. It is a typical approach to predict the next location of objects through mobility sequence tree generation by pattern mining the mobility of the object (Pfoser et al. 2000 ; Ying et al. 2011) [1] [2]. A further approach suggests the mobility tree which is expanded gradually by pattern mining (Gorawski and Jureczek. 2010) [3], and there is utilization with Location Based Service(LBS) to generate mobility patterns (Lee et al. 2004) [4]. Through such various methods, including mobility pattern tree generation and next location prediction, the combination of the probabilistic approach and data mining techniques has recently been

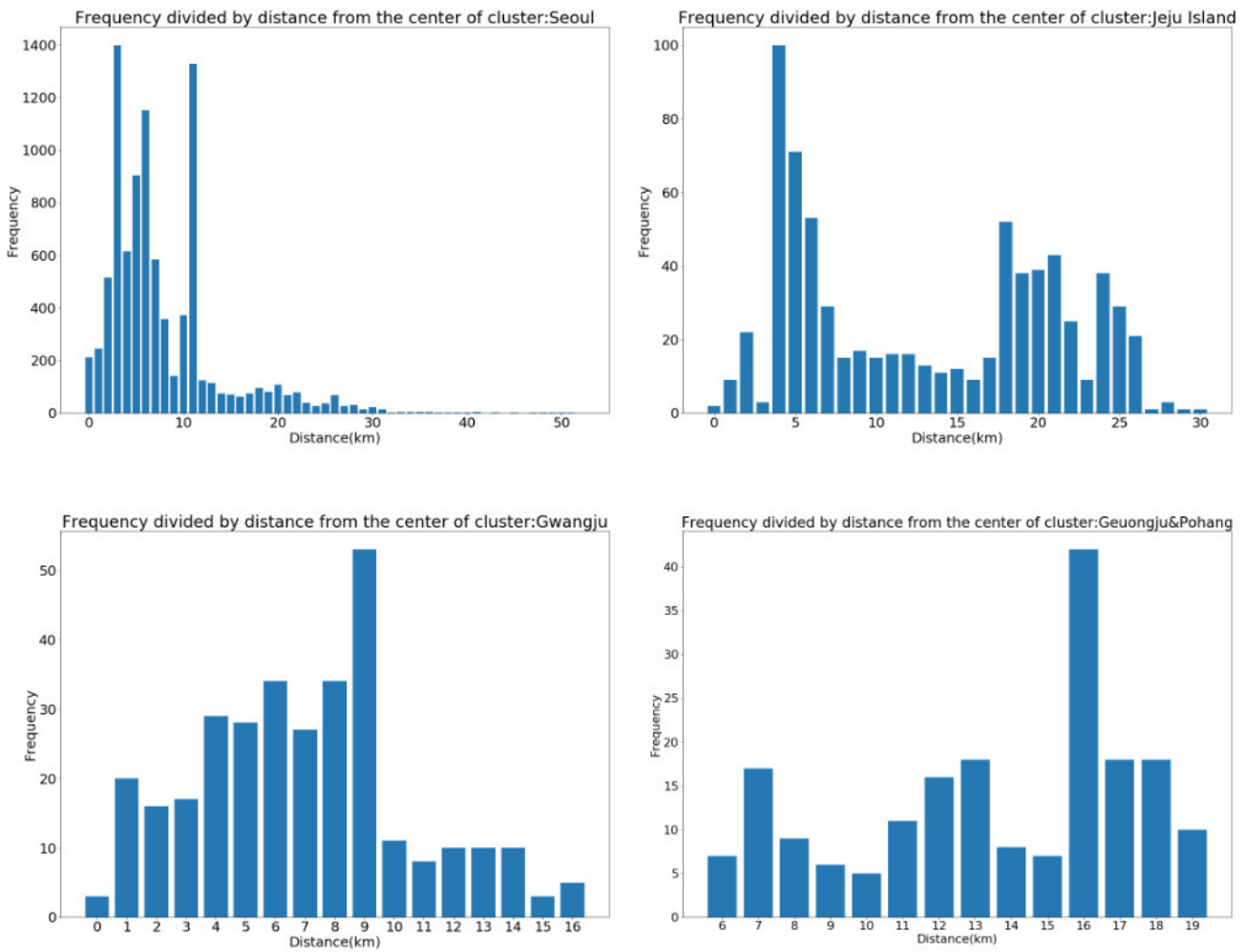


Fig. 1. Location Distribution inside Clusters

applied (Monreale et al. 2009 ; Jeung et al. 2008) [5] [6]. The graphs generated from the data mining results of accumulated trajectory patterns predict the next location (Morzy. 2006 ; Morzy. 2007) [7] [8]. Trajectory pattern can be utilized to mining past trajectories without trees, graphs or probability models (Giannotti et al. 2007) [9]. The success of combination of frequent trajectory and mobility rule for location prediction (Nhan and Ryu. 2006, Song and Choi. 2015) [10] [11] assumes diverse situations, such as disaster from the perspective of location prediction(Sudo et al. 2016) [12]. Markov chain-based approach has been applied on human mobility pattern and achieved remarkable performance on the next location prediction based on human mobility pattern(Baratchi et al. 2014) [13]. In the previous period, researchers have applied the data mining, trajectory pattern tree and Markov chain tools mainly.

From the view of GAN, GAN has achieved impressive outcomes in image generation (Alec Radford, Luke Metz, and Soumith Chintala. 2016) [14], image translation (Yunjey Choi

et al. 2018) [15], and there have been several recent researches to analysis synthetic data generation. Research regarding the activity patterns of neurons using GAN to generate the data can be found in (Molano-Mazon et al. 2018) [16]. Furthermore, movement trajectories based on socially acceptable behavior had been investigated as shown in (Gupta et al. 2018) [17]. Behavior includes passing or meeting of people during walking with parameters such as speed and direction. Current researchers for deep data generation on human mobility are (Alzantot, Chakraborty, and Srivastava. 2017) [18], using an architecture similar with GAN model, which is mixture of density networks to generate the acceleration time series data. However, this research did not use complete GAN architecture, and the input dataset of this research was not geopositioning data of trajectories but solely acceleration data. The purpose of our research is to generate novel mobility routes based on a plenty of daily mobility patterns.

TABLE I
BASIC PROPERTIES OF LOCATION CLUSTERS

Cluster No.	Location	Center of Cluster(Latitude / Longitude)	Total Number of Data in Cluster
Cluster 0	Seoul	33.48852999 126.47904523	9668
Cluster 1	Jeju Island	33.44230848 126.52364544	728
Cluster 2	Gwangju	35.16296341 126.88780208	317
Cluster 3	KyeongJu, Pohang	35.93293456 129.31300085	191

III. GEOPOSITIONING DATA ANALYSIS

The mobility data collected by positioning devices, such as smart phones, have the latitude, longitude and time information. The raw data we use are containing the location information of a specific object for more than three years.

A. Macroscopic View

In analyzing data from a macro perspective, the K-means clustering method is applied to extract largely four clusters based on location and density from the collected data distribution [19]. The basic properties of each location cluster are shown in table I. Fig. 1 shows histograms of the location points of each cluster. The frequency values are shown according to the distance between a center and positioning data in each location cluster [20]. Cluster 0 contains a relatively large number of location data, and this cluster reflects the location data in lifestyle. On the other hand, a low number of data is collected at cluster 1, 2, 3, and its result shows the irregular mobility pattern of object, such as a trip. This implies that the micro-mobility of an object indicates remarkable differences in each cluster depending on the purpose to visit.

B. Microscopic View

Understanding correlation between visited points is a starting point for expanding correlation between day-to-day movement trajectories. In cluster 0, there are mobility patterns on the movement of lifestyle, and the mobility patterns consist of visited points in object's daily life. To analyze correlation among distinguished points in cluster 0, it is necessary to extract past visited points. Distinguished points can be found applying K-means clustering algorithm to each day of the three-year full mobility data. Redundancies of 185 geographical points were eliminated, and the entire mobility pattern in cluster 0, made up of each major point, consists of about 2,700 distinguished points.

Learning connection between distinguished points and connection between each movement pattern, where the points are clustered, is identifying the mobility tendency of an object. Our GAN-based model will put daily movement patterns in Discriminator as input dataset, recognizing the existing patterns as real data, and will train dozens of times for Generator to generate mobility route similar to input dataset. This begins with the notion that a mobility route will be created based on object's mobility tendency.

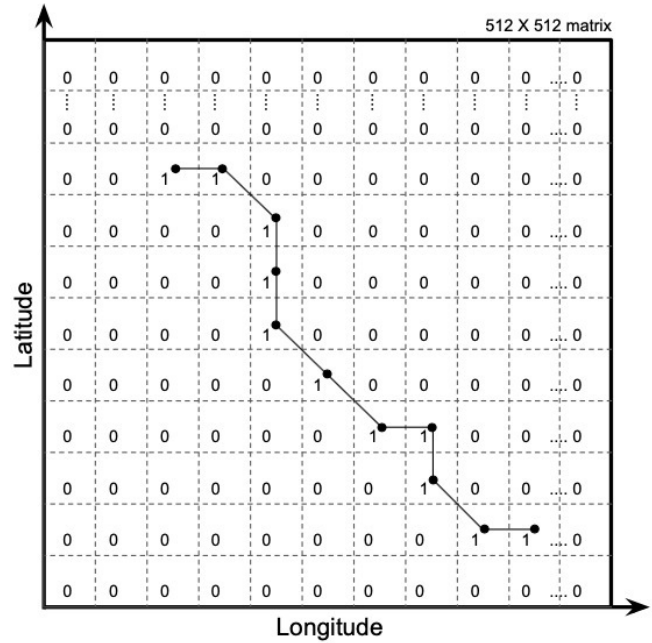


Fig. 2. Area Partitioning and mapping

IV. GEOPOSITIONING DATA PROCESS

A. Data Pre-processing

As the wide range of location dataset is somewhat excessive for GAN model to train, proper data preprocessing is required. The unstructured raw data is initially informal structure, and could not apply to GAN model directly. The area is limited to the size of 3 kilometer in vertical length and 3 kilometer in horizontal length. The input dataset was structured using area partitioning method. The mobility patterns were partitioned as shown in Fig. 2. As a practical approach, area partitioning method can represent daily movement patterns as matrix. In detail, each visited area is mapped from intersection between the trajectory mobility pattern and unit areas. Each unit area size is about 0.06 square kilometer. By the characteristic of the ge positioning data, positioning error, such as value of latitude and longitude, is effectively found from the fifth decimal point of values. That is, eliminating the digits from the 6th to 11th decimal places do not distort the location information. We applied the round function through the area partitioning method. To preserve all positioning information in matrices, the CNN is utilized with 4 layers with each filter size of

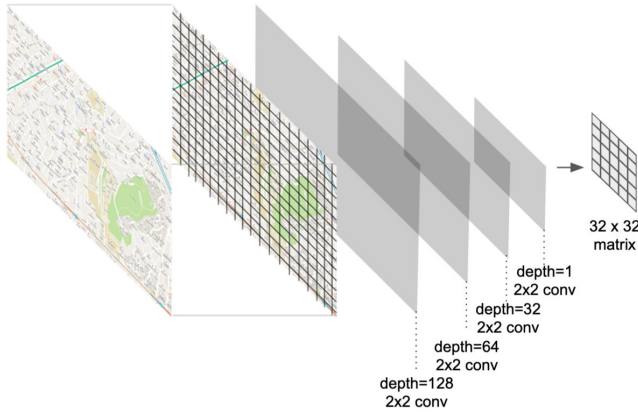


Fig. 3. Process of Convolution

2x2x128, 2x2x64, 2x2x32, and 2x2x1. In data convolution process, input data for GAN model are transformed to 32x32 sized matrices. This problem is solved by making the mobility data to pass multiple CNN layers with Leaky ReLU as an activation function which is convolution stage. Fig. 3 shows the process of convolution. The volume of input dataset was not enough to train the mobility patterns by our model, in detail, the individual accumulated daily mobility pattern. To improve this problem, synthetic inputs, which are extracted by randomly shuffling of 40% to 90% of input data, are duplicated so that GAN model can train various sets of input data. Through this augmentation method, insufficient input dataset amplified 100 times.

B. Data Post-processing

The output of our GAN model is 32x32 sized matrices. In order to visualize output data on the map, deconvolution as post-processing is required. The CNN as deconvolution layers is used with 4 layers. In this deconvolution process, Nearest Neighbor function is utilized to resize process which transforms the output to 64x64, 128x128, 256x256 and 512x512 size step by step. Fig. 4 shows the steps of deconvolution toward visualization on geographical maps.

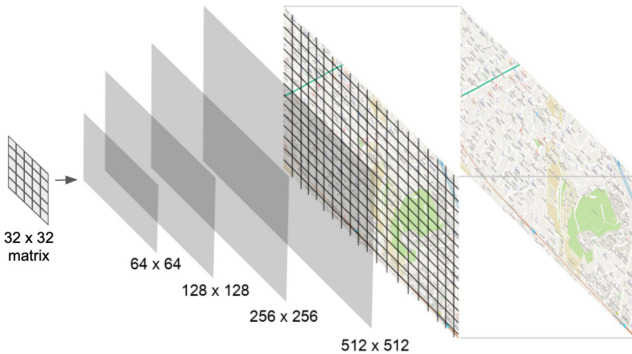


Fig. 4. Process of Deconvolution

V. METHODOLOGY WITH GAN

GAN was firstly introduced from (Goodfellow et al. 2014) [21], which consists of a Generator and a Discriminator, structured by neural networks. The basic theory of GAN is started below the Non-Saturating game. In detail, the Discriminator minimizes the value of equation 1 and the Generator maximizes the value of equation 2. Generally, GAN model has the saturating issue, which occurs with initialization. Equation 3 will be used as an objective function to solve this issue (Fedus et al. 2017) [22].

$$\max_D \min_G V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(x)} [\log(1 - D(G(z)))] \quad (1)$$

$$\max_D \min_G V(D, G) = E_{z \sim p_z(x)} [\log(1 - D(G(z)))] \quad (2)$$

$$\min_G E_{z \sim p_z(x)} [\log(1 - D(G(z)))] = \max_G E_{z \sim p_z(x)} [\log D(G(z))] \quad (3)$$

We proposed GAN to generate diverse novel mobility routes. In this GAN, the two neural networks compete each other to improve ability to generate mobility route with latitude and longitude features. Discriminator tries to discriminate the real mobility route, training from the accumulated movement trajectory of object and exporting the probability value which is close to “1”, when the input data is distinguished as real data. On the other hand, the Generator has the random latent vector as input, and generates a matrix. Discriminator tries to compare each matrix generated from Generator with input dataset. When it is indistinguishable if the matrix is from Generator or input dataset, the Generator can suggest a novel route which put together a wide range of daily route. Fig. 6 and Fig. 7 shows our GAN architecture for experiment.

VI. EXPERIMENT

A. Test of Convolution and Deconvolution

Despite the data processing, we could find that the geopositioning data are preserved when the data are passed through convolution and deconvolution layer. Fig. 8 shows the original route, convolutionized route and deconvolutionized route as diagram. In test of convolution and deconvolution, fig. 9-(a) shows mapped route from raw data, and fig. 9-(b) shows route after deconvolution process without error. From this result, it is appropriate to apply convolution and deconvolution layer with GAN model.

B. Experiment for Activation Function

In order to figure out the better result, generating route, we experimented with varying activation functions for Generator and Discriminator. Seven activation functions are applied in this work, such as ReLU, ReLU6, Softplus, Tanh, Softsign and eLU (Fig. 5) with 20,000 and 50,000 epoch. Among

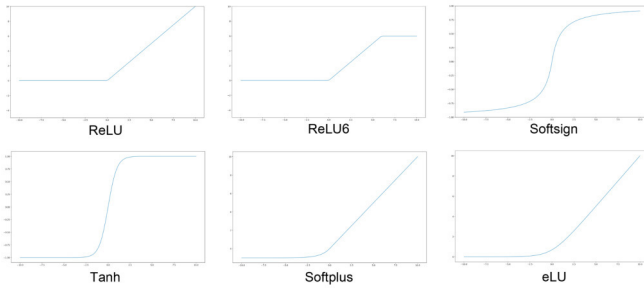


Fig. 5. Activation Functions

the result with the 7 activation functions, ReLU activation function made the worst output, and Tanh activation function made the best output. During training with ReLU, it has possibility to partially miss some positioning data. On the other hand, during training with Tanh, even if Tanh reduce the positioning data, the mobility route can be produced without missing positioning data. This is because the network of the model tends to partially lose its path data as it passes through the ReLU activation function while learning with the input paths. On the other hand, because Tanh minimizes the path as it passes the function and does not make it zero, the path consisting of frequency can be created without missing value. Fig. 10 shows the result with ReLU activation function for 20,000 and 50,000 epochs. Comparing between Fig. 10(a) and Fig. 10(b), ReLU formed accumulated path group. However, ReLU generates clusters rather than generating paths even if epoch increases. Fig. 11 shows the results of two epoch processes with 20,000 and 50,000 epochs with Tanh activation function. About 20,000 epochs, ambiguous paths are created. Approximately more than 50,000 epochs, trained Generator creates a new path. Since the inputs of GAN model are diverse, the outputs seem to be slightly different but suitable paths.

C. Unexplored Route

Fig. 12 shows two routes generated by GAN. Fig. 12-(a) shows typical output generated. Fig. 12-(b) is not trivial output comparing to output shown in Fig. 10-(a). GAN generated normal routes and it also generated additional routes. That is GAN creates unexplored, new route.

VII. CONCLUSION

We developed a method to generate human mobility routes based on Generative Adversarial Networks (GANs) which are specialized in image generation. We successfully trained GAN model on individual mobility data, and the model could suggest novel routes in about 3 square kilometer of area, reflecting preference of specific object. In addition, we show the possibility to train any extensive trajectory pattern instead of massive geopositioning data by applying CNN layers in front of the model. The essence of our research is that GAN model can train on the continuous sequential data, such as trajectory pattern and can generate mobility route with additional information. The purpose of this model is

to provide unexplored routes which is based on an object's daily mobile data. By training an object's mobile data, we could generate mobility routes, which the users are likely to explore in their future. Therefore, generated mobility routes are fundamental basis to individual recommendation system. By combining our model with Location Based Service (LBS) and Recommender Systems will allow us to provide whole new service to users. Users will be recommended places such as personalized restaurants, shops, and other utilities within unexplored generated mobility routes.

While the better purpose in our experiment is to create mobile route on random location, the model's ability to generate the route would be bounded with the high randomness. Additionally, we did not interconnect between mobility data and human behavior factor. Further, the biased mobility data, which is adhere to specific location such as school or workplace, caused drawbacks in the model training, and the application based on geolocation system has inherent error. As we mentioned the combination of human behavior factor such as hobby, job, personality, etc., the research that combines with personal factor with our method would be an advanced method to generate mobility route more connected with mobility tendency, offering the personal customized solution. In future work, we aim to apply social media service which includes personal behavior features. It would be possible to identify the purpose of movement and specific time period. This can be expanded from individual to group with similar personality. As a result, it would be possible to create mobility route which is more customer-oriented.

REFERENCES

- [1] D. Pfoser, C. S. Jensen, and Y. Theodoridis, "Novel approaches in query processing for moving object trajectories," in *Proceedings of the 26th International Conference on Very Large Data Bases*, ser. VLDB '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000. ISBN 1-55860-715-3 pp. 395–406. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645926.672019>
- [2] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng, "Semantic trajectory mining for location prediction," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '11. New York, NY, USA: ACM, 2011. doi: 10.1145/2093973.2093980. ISBN 978-1-4503-1031-4 pp. 34–43. [Online]. Available: <http://doi.acm.org/10.1145/2093973.2093980>
- [3] M. Gorawski and P. Jureczek, "Continuous pattern mining using the fcpgrowth algorithm in trajectory data warehouses," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, pp. 187–195. [Online]. Available: https://doi.org/10.1007%2F978-3-642-13769-3_23
- [4] J. W. Lee, O. H. Paek, and K. H. Ryu, "Temporal moving pattern mining for location-based service," *Journal of Systems and Software*, vol. 73, no. 3, pp. 481–490, nov 2004. doi: 10.1016/j.jss.2003.09.021. [Online]. Available: <https://doi.org/10.1016%2Fj.jss.2003.09.021>
- [5] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009. doi: 10.1145/1557019.1557091 pp. 637–646.
- [6] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *2008 IEEE 24th International Conference on Data Engineering*. IEEE, apr 2008. doi: 10.1109/icde.2008.4497415. [Online]. Available: <https://doi.org/10.1109%2Ficde.2008.4497415>
- [7] M. Morzy, "Prediction of moving object location based on frequent trajectories," in *Computer and Information Sciences ISCIS 2006*. Springer Berlin Heidelberg, 2006, pp. 583–592. [Online]. Available: https://doi.org/10.1007%2F11902140_62

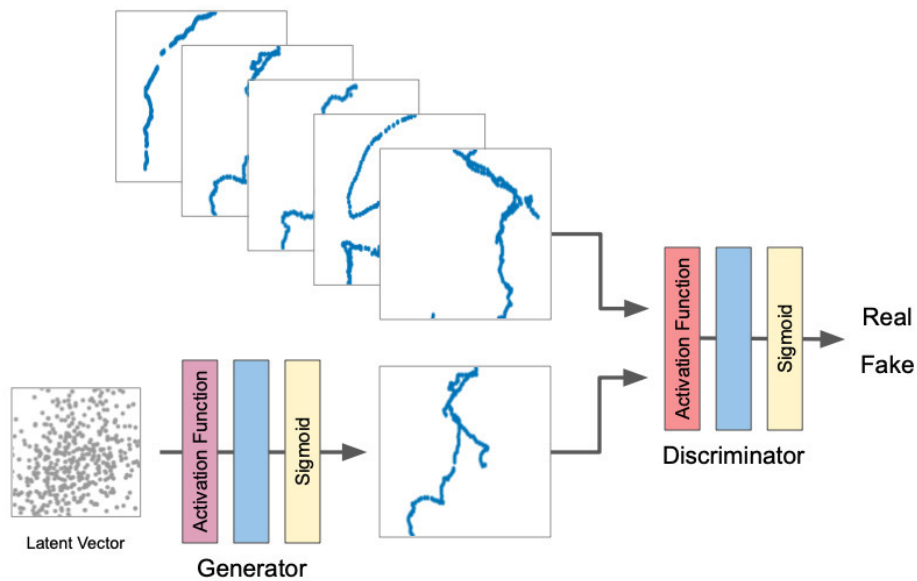


Fig. 6. Architecture for our GAN

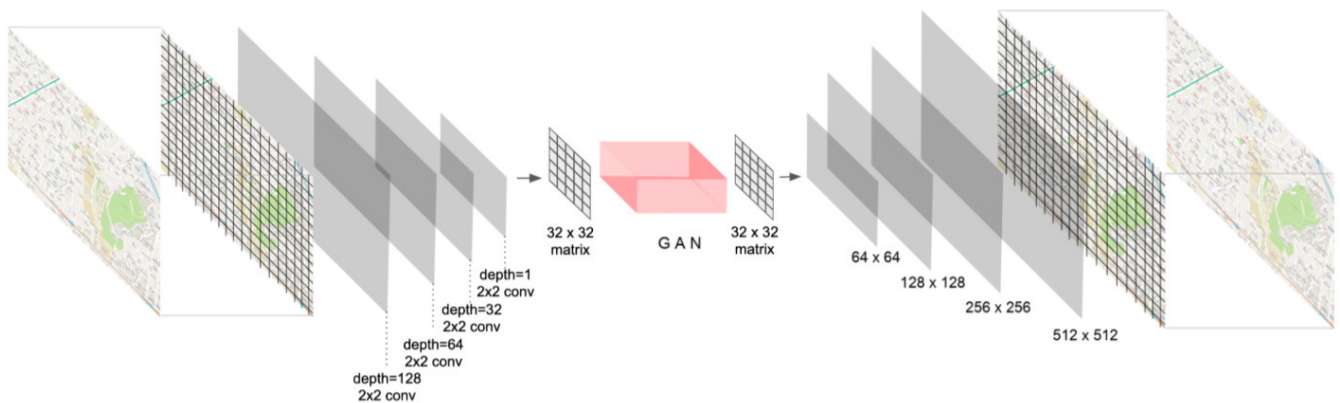


Fig. 7. Whole Architecture of our Model

- [8] —, “Mining frequent trajectories of moving objects for location prediction,” in *Machine Learning and Data Mining in Pattern Recognition*. Springer Berlin Heidelberg, 2007, pp. 667–680. [Online]. Available: https://doi.org/10.1007%2F978-3-540-73499-4_50
- [9] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, “Trajectory pattern mining,” in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007. doi: 10.1145/1281192.1281230 pp. 330–339.
- [10] V. T. H. Nhan and K. H. Ryu, “Future location prediction of moving objects based on movement rules,” in *Intelligent Control and Automation*. Springer Berlin Heidelberg, 2006, pp. 875–881. [Online]. Available: https://doi.org/10.1007%2F11816492_112
- [11] H. Y. Song and D. Y. Choi, “Defining measures for location visiting preference,” *Procedia Computer Science*, vol. 63, pp. 142–147, 2015. doi: 10.1016/j.procs.2015.08.324. [Online]. Available: <https://doi.org/10.1016%2Fj.procs.2015.08.324>
- [12] A. Sudo, T. Kashiyama, T. Yabe, H. Kanasugi, X. Song, T. Higuchi, S. Nakano, M. Saito, and Y. Sekimoto, “Particle filter for real-time human mobility prediction following unprecedented disaster,” in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. SIGSPACIAL '16. New York, NY, USA: ACM, 2016. doi: 10.1145/2996913.2997000. ISBN 978-1-4503-4589-7 pp. 5:1–5:10. [Online]. Available: <http://doi.acm.org/10.1145/2996913.2997000>
- [13] M. Baratchi, N. Meratnia, P. J. M. Havinga, A. K. Skidmore, and B. A. K. G. Toxopeus, “A hierarchical hidden semi-markov model for modeling mobility data,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp 14 Adjunct*. ACM Press, 2014. doi: 10.1145/2632048.2636068. [Online]. Available: <https://doi.org/10.1145%2F2632048.2636068>
- [14] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434v2*, 2016.
- [15] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018. doi: 10.1109/cvpr.2018.00916. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2018.00916>
- [16] M. Molano-Mazon, A. Onken, E. Piasini, and S. Panzeri, “Synthesizing

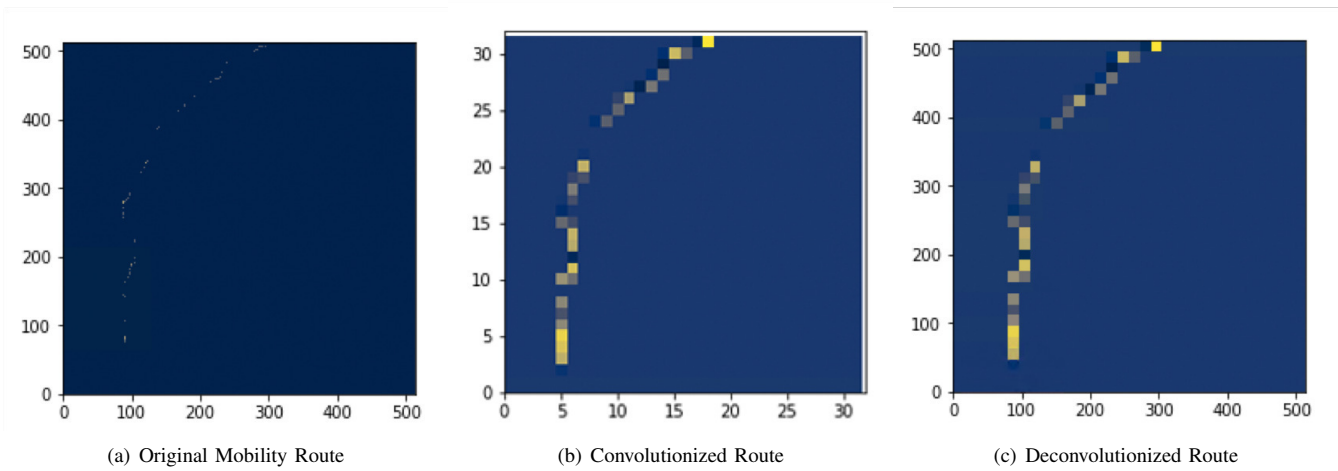


Fig. 8. Result of Convolution and Deconvolution

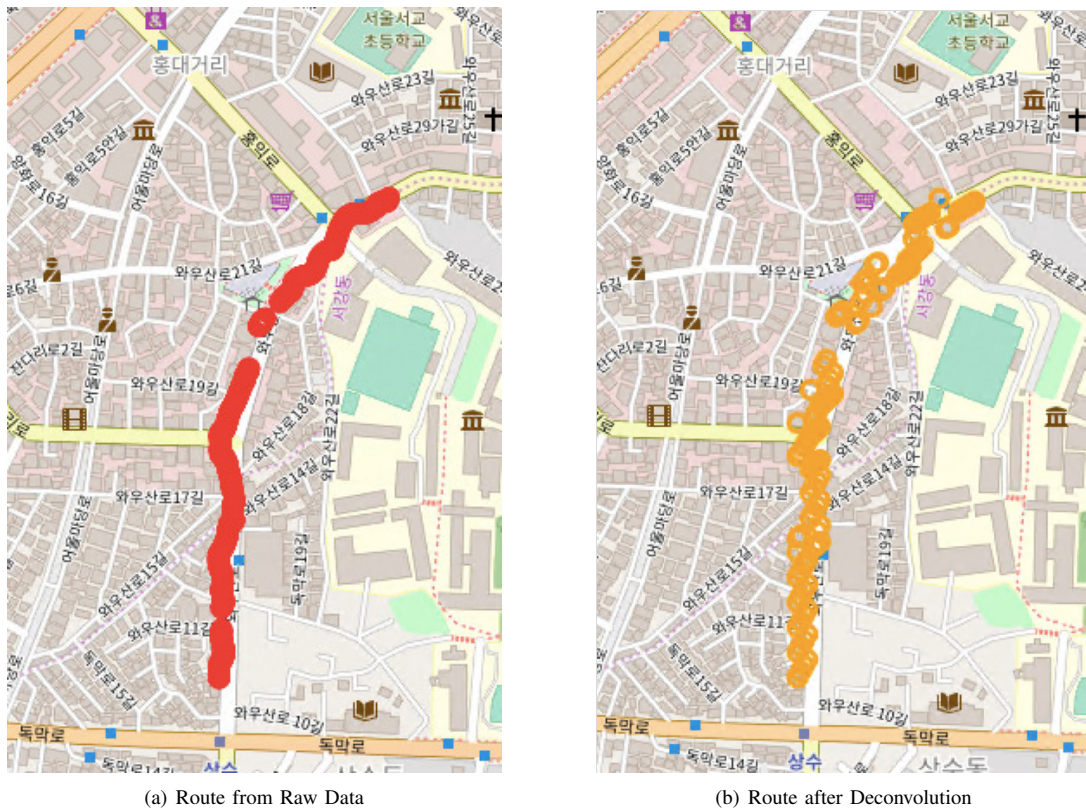


Fig. 9. Mapped Result from Raw Mobility Data versus after Deconvolution

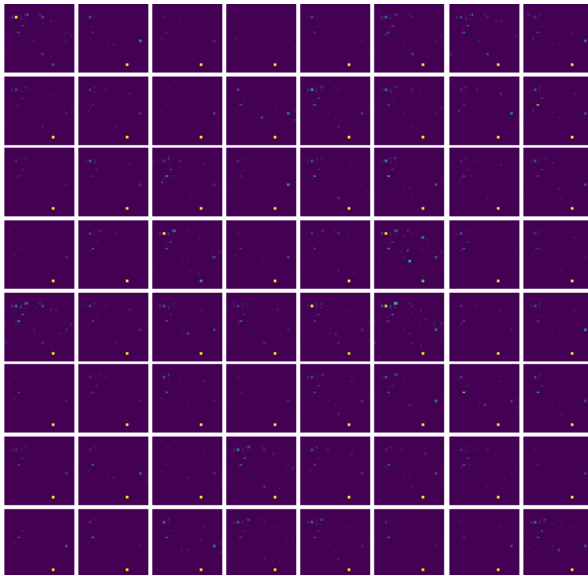
realistic neural population activity patterns using generative adversarial networks,” *arXiv preprint arXiv:1803.00338*, 2018.

[17] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018. doi: 10.1109/cvpr.2018.00240. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2018.00240>

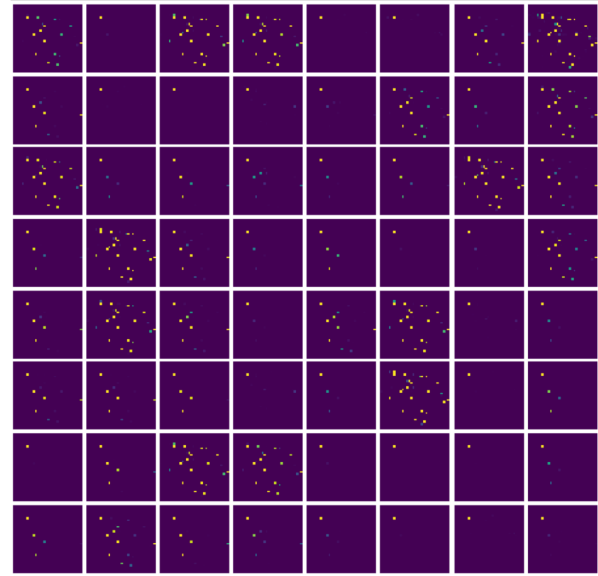
[18] M. Alzantot, S. Chakraborty, and M. Srivastava, “Sensegen: A deep learning architecture for synthetic sensor data generation,” in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, Mar 2017. doi: 10.1109/percomw.2017.7917555. [Online]. Available: <https://doi.org/10.1109%2Fpercomw.2017.7917555>

[19] S. Na, L. Xumin, and G. Yong, “Research on k-means clustering algorithm: An improved k-means clustering algorithm,” in *2010 Third International Symposium on intelligent information technology and security informatics*. IEEE, 2010. doi: 10.1109/IITSI.2010.74 pp. 63–67.

[20] P. Reinecke, T. Krauss, and K. Wolter, “Hyperstar: Phase-type fitting made easy,” in *2012 Ninth International Conference on Quantitative*

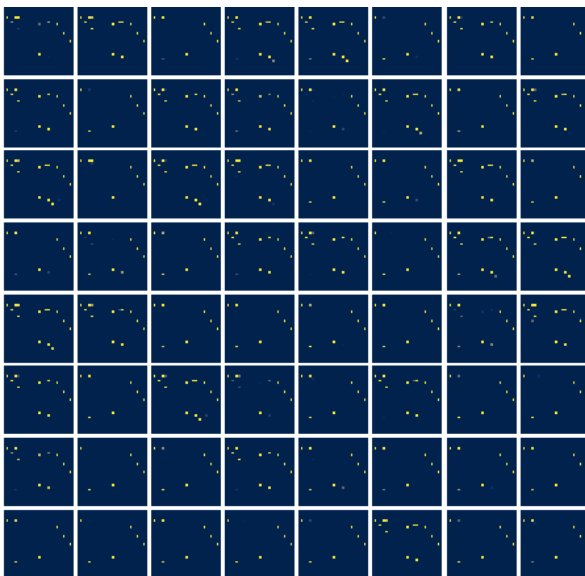


(a) 20,000 epochs with ReLU

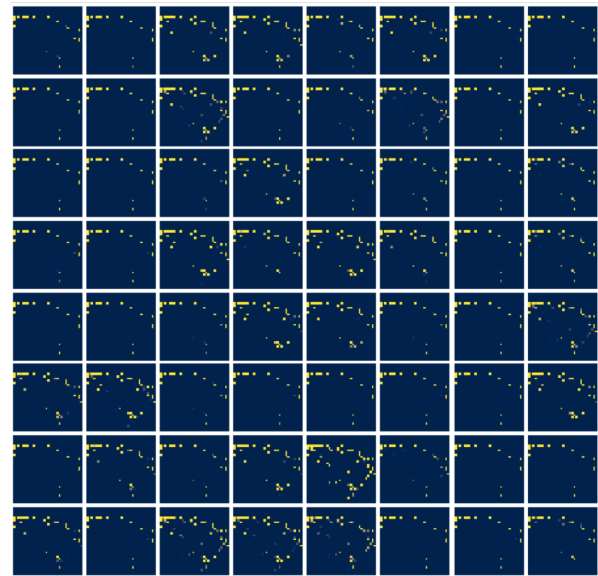


(b) 50,000 epochs with ReLU

Fig. 10. Result with ReLU activation function



(a) 20,000 epochs with Tanh



(b) 50,000 epochs with Tanh

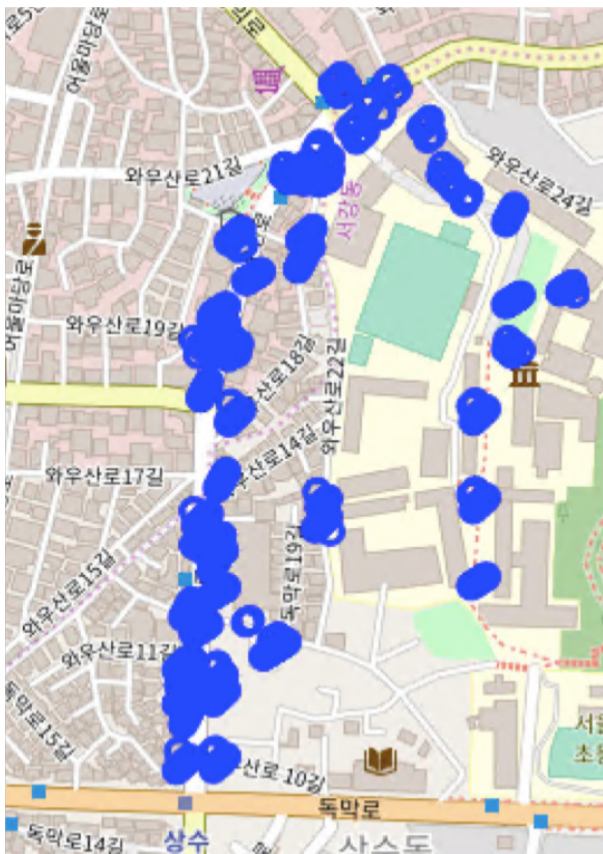
Fig. 11. Result with Tanh activation function

Evaluation of Systems. IEEE, Sep 2012. doi: 10.1109/qest.2012.29. [Online]. Available: <https://doi.org/10.1109/qest.2012.29>

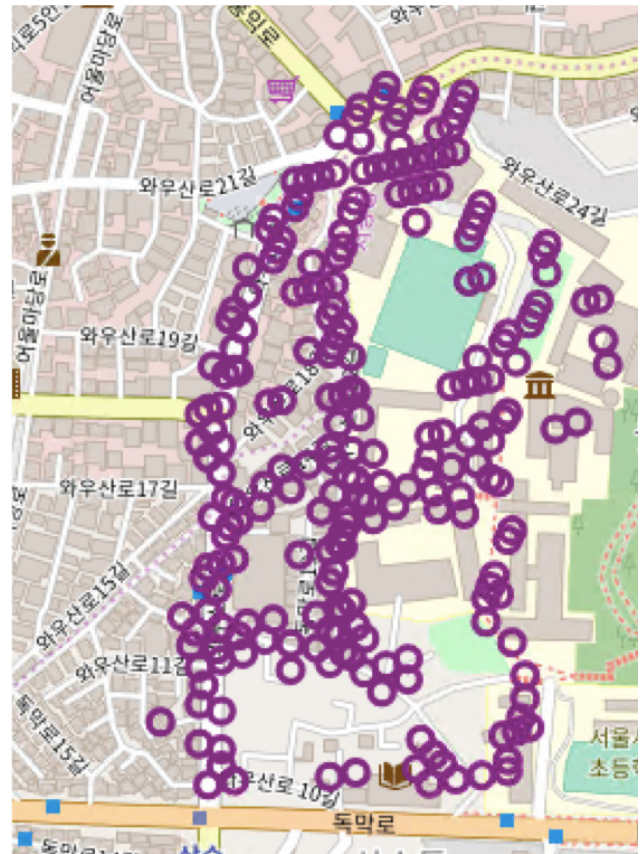
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds.

Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>

- [22] W. Fedus, M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed, and I. Goodfellow, "Many paths to equilibrium: Gans do not need to decrease adivergence at every step," 10 2017.



(a) General Generated Route



(b) Generated Route with Additional Information

Fig. 12. Route Generated by GAN