Generating Urban Mobility Data Sets Using Scalable GANs

Abhinav Jauhri & John Paul Shen
ECE Department
Carnegie Mellon University

{ajauhri, jpshen}@cmu.edu

Mobility21
A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING

Objective

Generate city-scale human mobility data using Generative Adversarial Networks (GANs) for Intelligent Transportation Systems.

Outline

- Motivation
- Spatial and Temporal Variations
- Generative Adversarial Networks (GANs)
- Ride Requests to Images
- Experiments
- Conclusions

Outline

- Motivation
- Spatial and Temporal Variations
- Generative Adversarial Networks (GANs)
- Ride Requests to Images
- Experiments
- Conclusions

Motivation

- Tackle challenges related to urban transportation in large cities --
  - How to perform pooling?
  - What are the savings of pooling passengers in terms of travel distance reduction, vehicle count reduction?
  - What are the savings of placing vehicles smartly?
  - Many more …

- Access of data for researchers and civic authorities to conduct experiments related to Intelligent Transportation Systems (ITS).

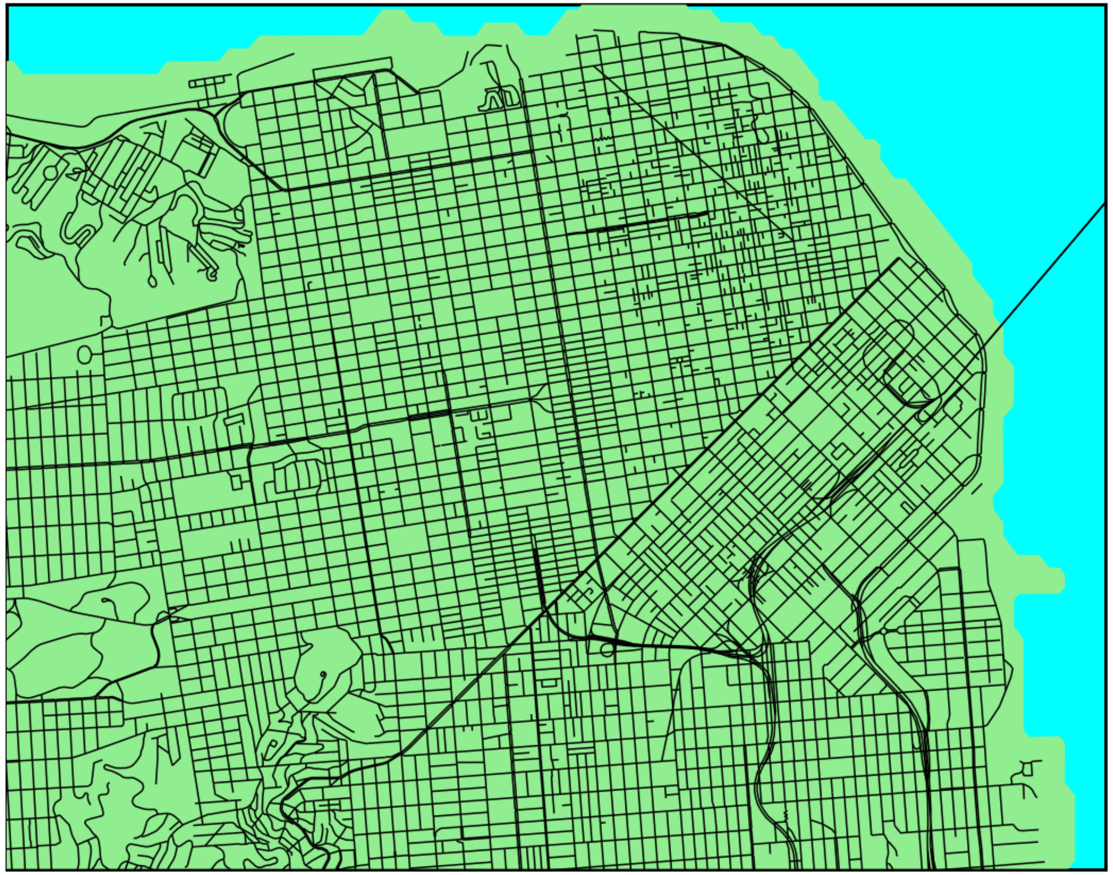
- Modeling challenge -- tackle a real-world problem using GANs

Outline

- Motivation
- Spatial and Temporal Variations
- Generative Adversarial Networks (GANs)
- Ride Requests to Images
- Experiments
- Conclusions

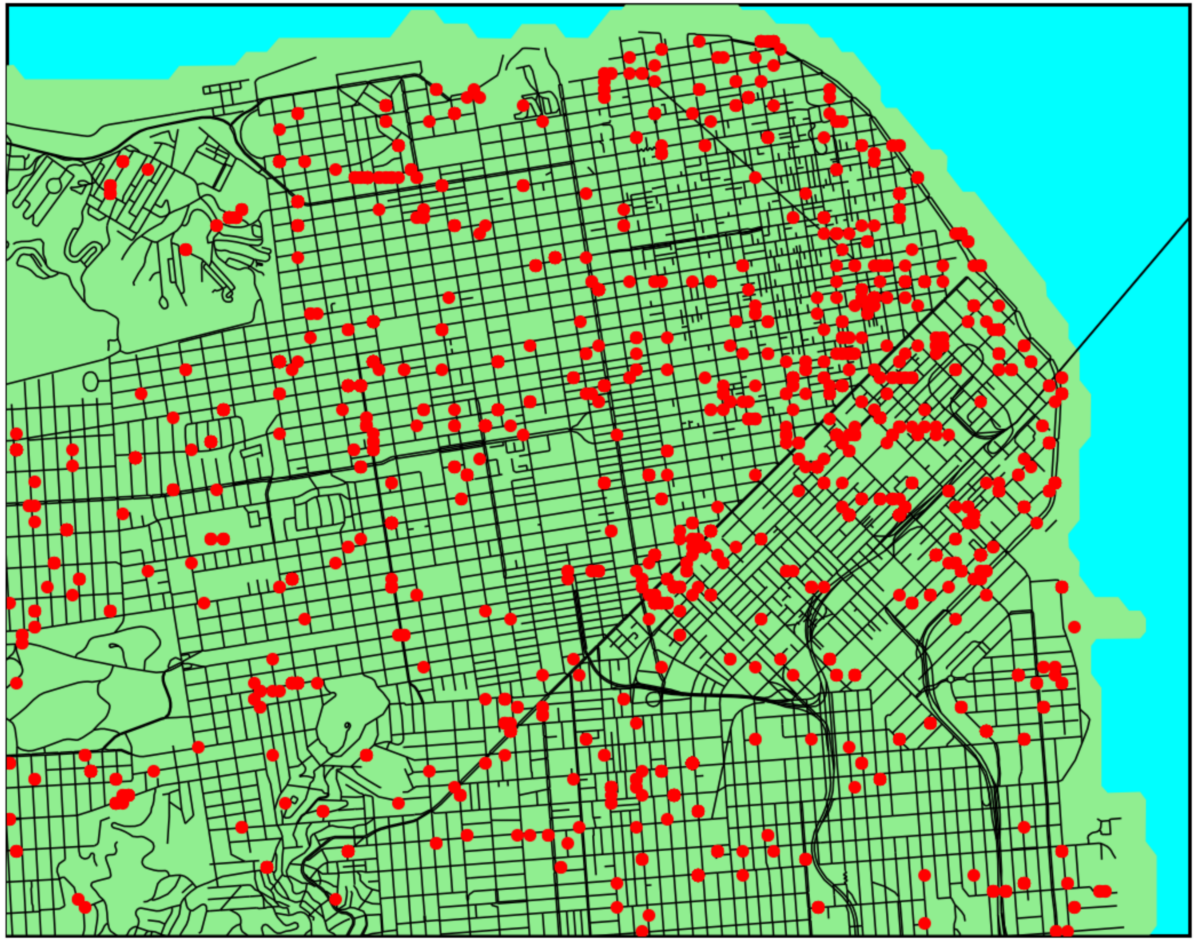
Distribution of pickup locations in San Francisco

Downtown San Francisco



Distribution of pickup locations in San Francisco

Downtown San Francisco

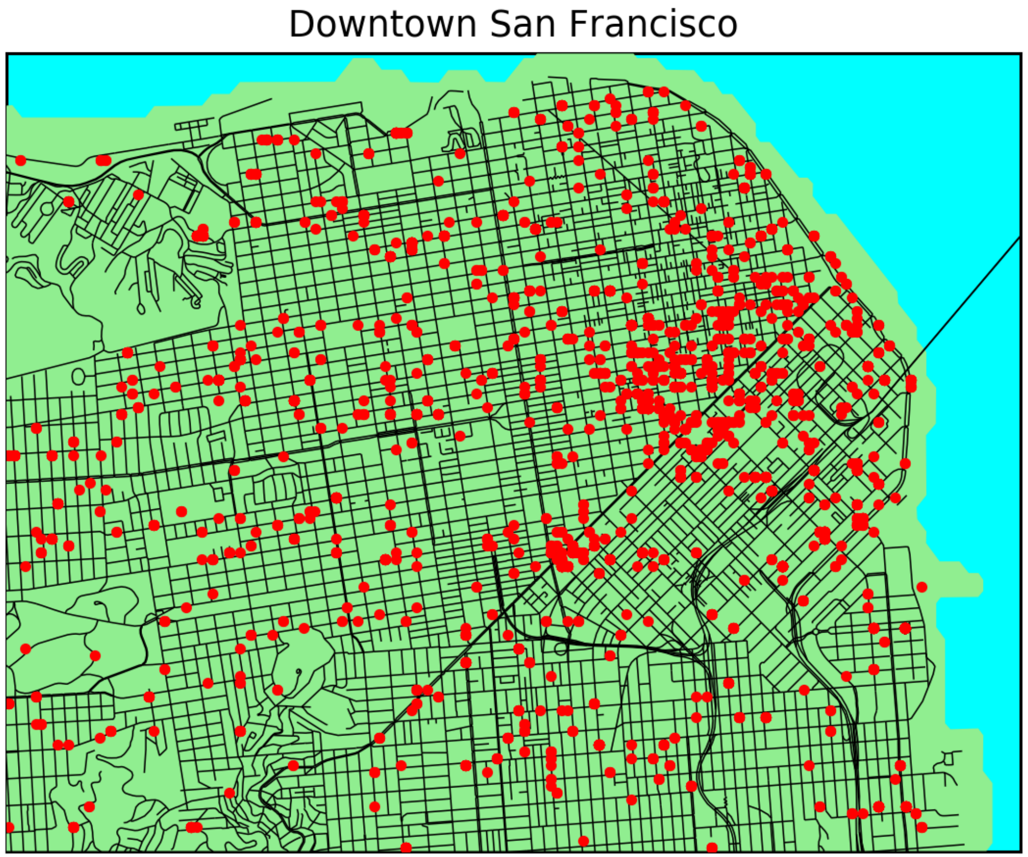


Every red dot represents the source of a ride request. Ride requests aggregated over a **5-minute time snapshot at 5pm**.

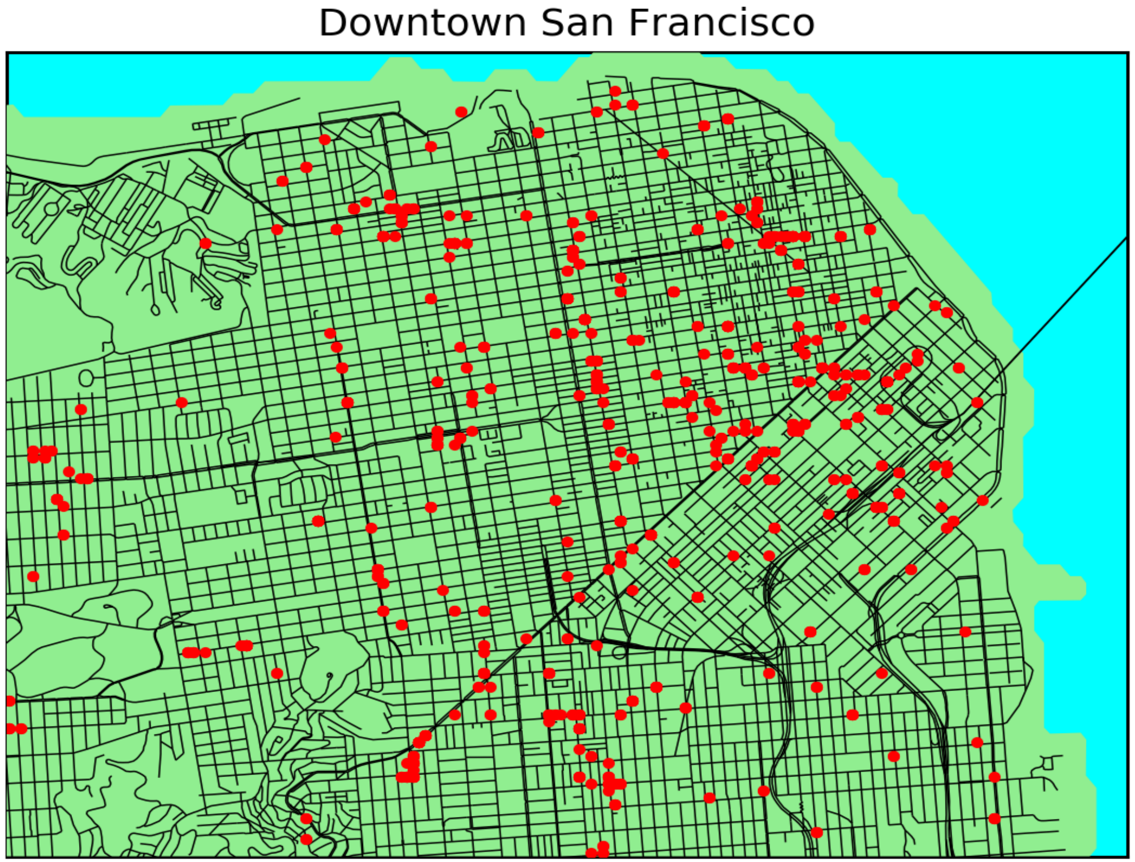
Mobility21 A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING

Distribution of pickup locations in San Francisco



Ride requests aggregated over a **5-minute time snapshot at 6pm**.

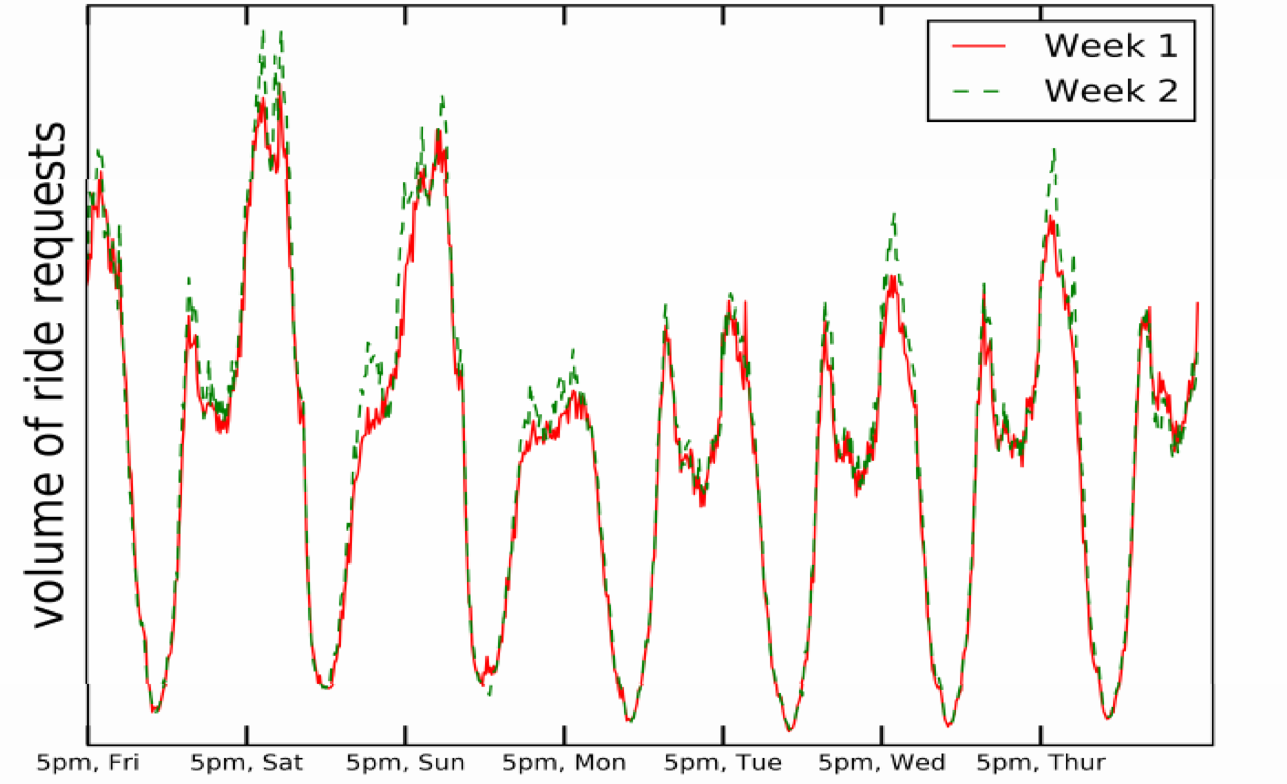


Ride requests aggregated over a **5-minute time snapshot at 2am**.

Mobility21
A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING

Volume of Ride Requests over a week



Quantity of ride requests for multiple weeks

Observation #1

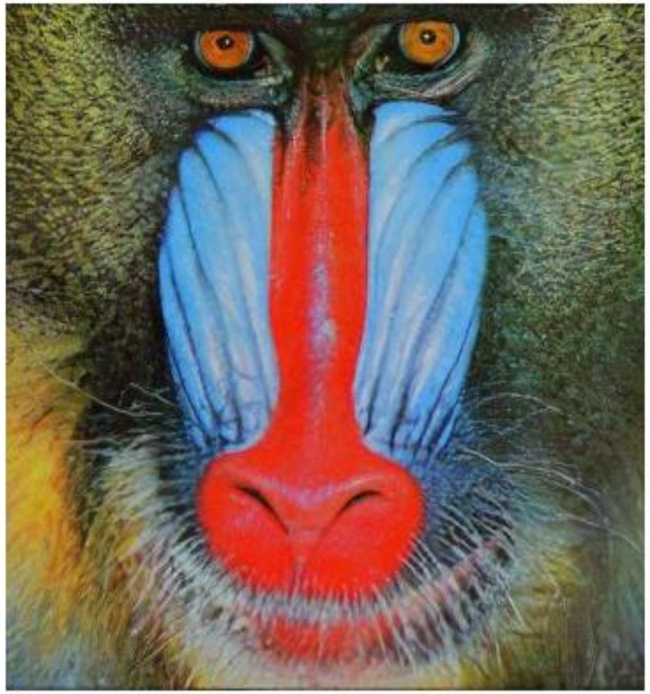
Human mobility patterns are highly dynamic both spatially and temporally.

Outline

- Motivation
- Spatial and Temporal Variations
- Generative Adversarial Networks (GANs)
- Ride Requests to Images
- Experiments
- Conclusions

Generative Adversarial Networks: Sample Generation

Generative Adversarial Networks: Image Super Resolution



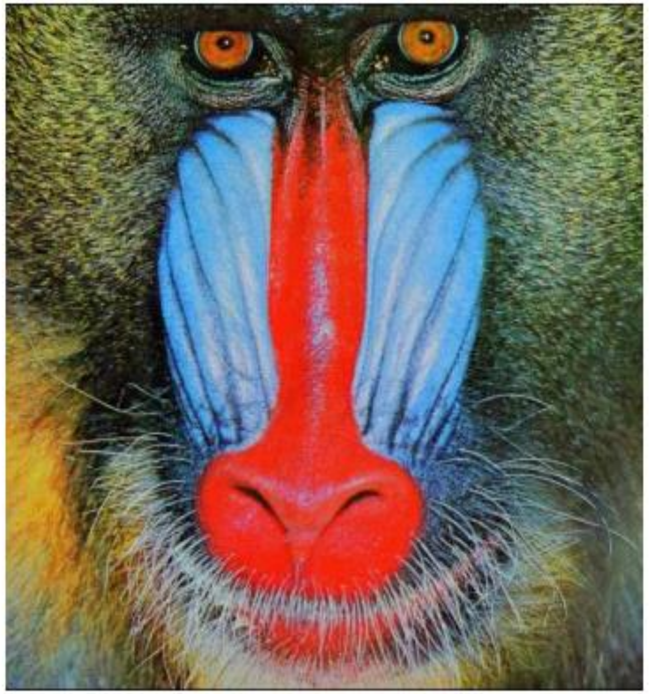


Image generated using GAN (left) is almost identical to the original (right) [Ledig et. al., CVPR 2017]

Generative Adversarial Networks: Image Inpainting

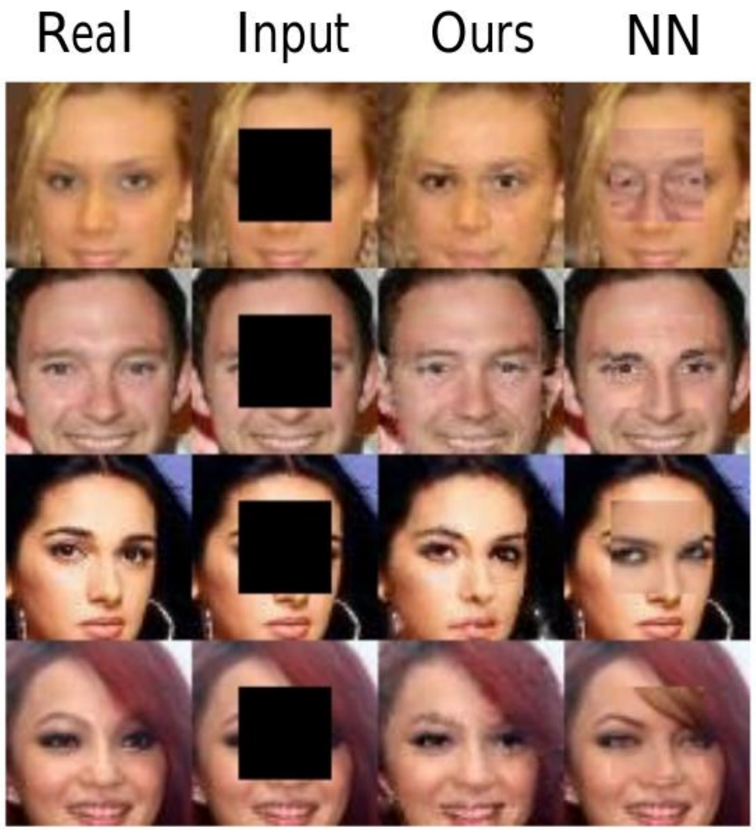
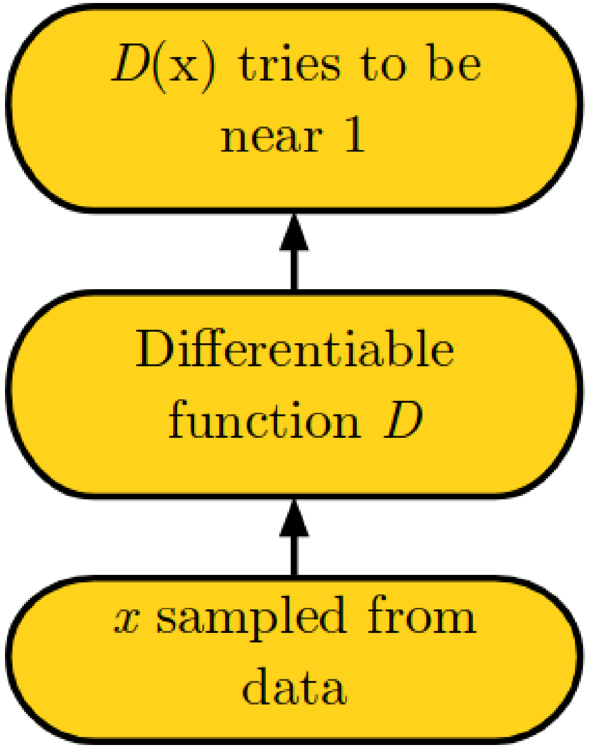


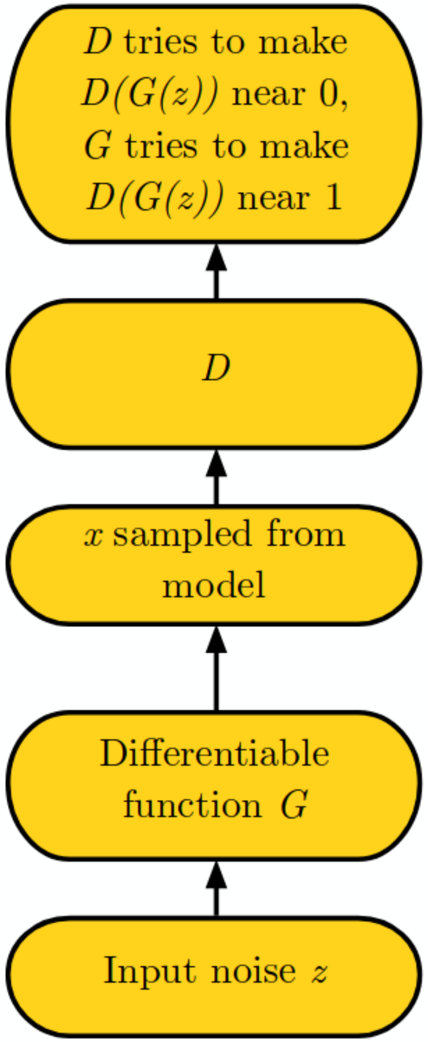
Image inpainting using GANs [Yeh et. al., CVPR 2017]

Mobility21
A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING

Generative Adversarial Networks: Framework





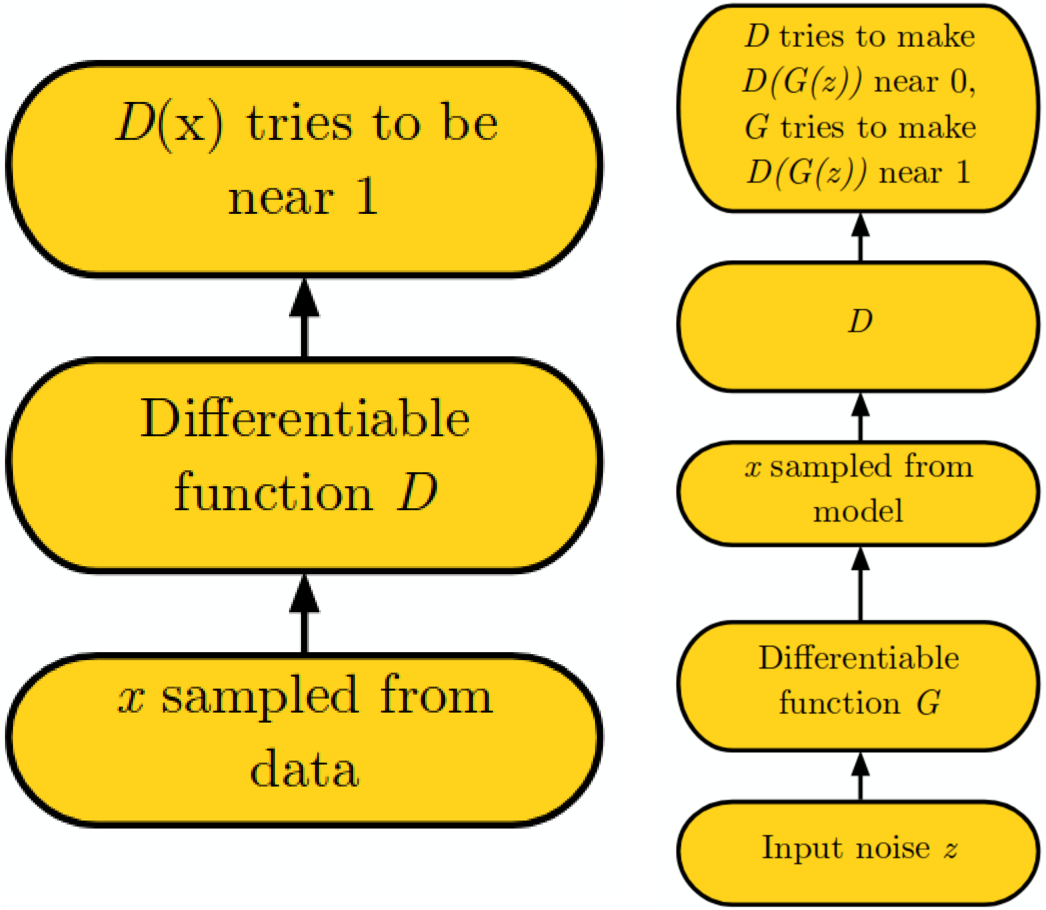
D - Discriminator
G - Generator

Image Source: Goodfellow et. al., 2014]

Mobility21 A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING

GANs for Human Mobility



Objective -- How to generate series of images for consecutive time steps representing human mobility data?

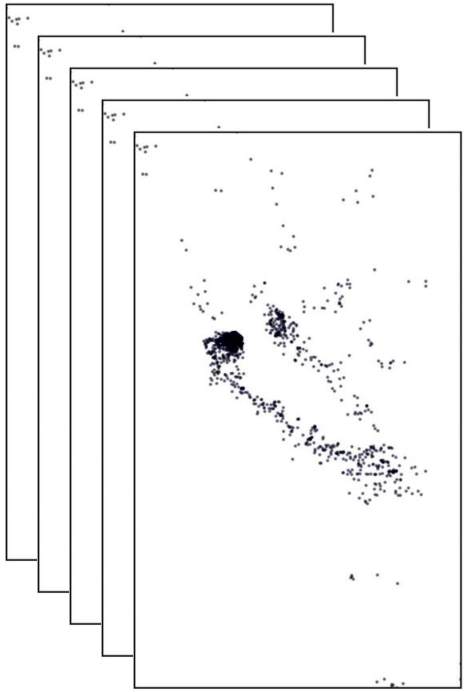


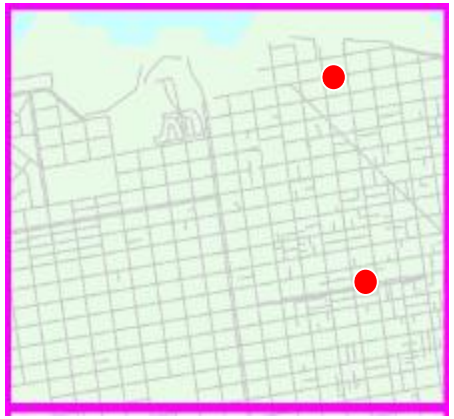
Image Source: Goodfellow et. al., 2014]

Outline

- Motivation
- Spatial and Temporal Variations
- Generative Adversarial Networks (GANs)
- Ride Requests to Images
- Experiments
- Conclusions

Ride Requests to Images

Each ride request's originating location is represented by a **<latitude, longitude>** on the geographical space. We discretize the map into 50x50 meters represented by a pixel.



Map with two ride requests shown with two black dots.

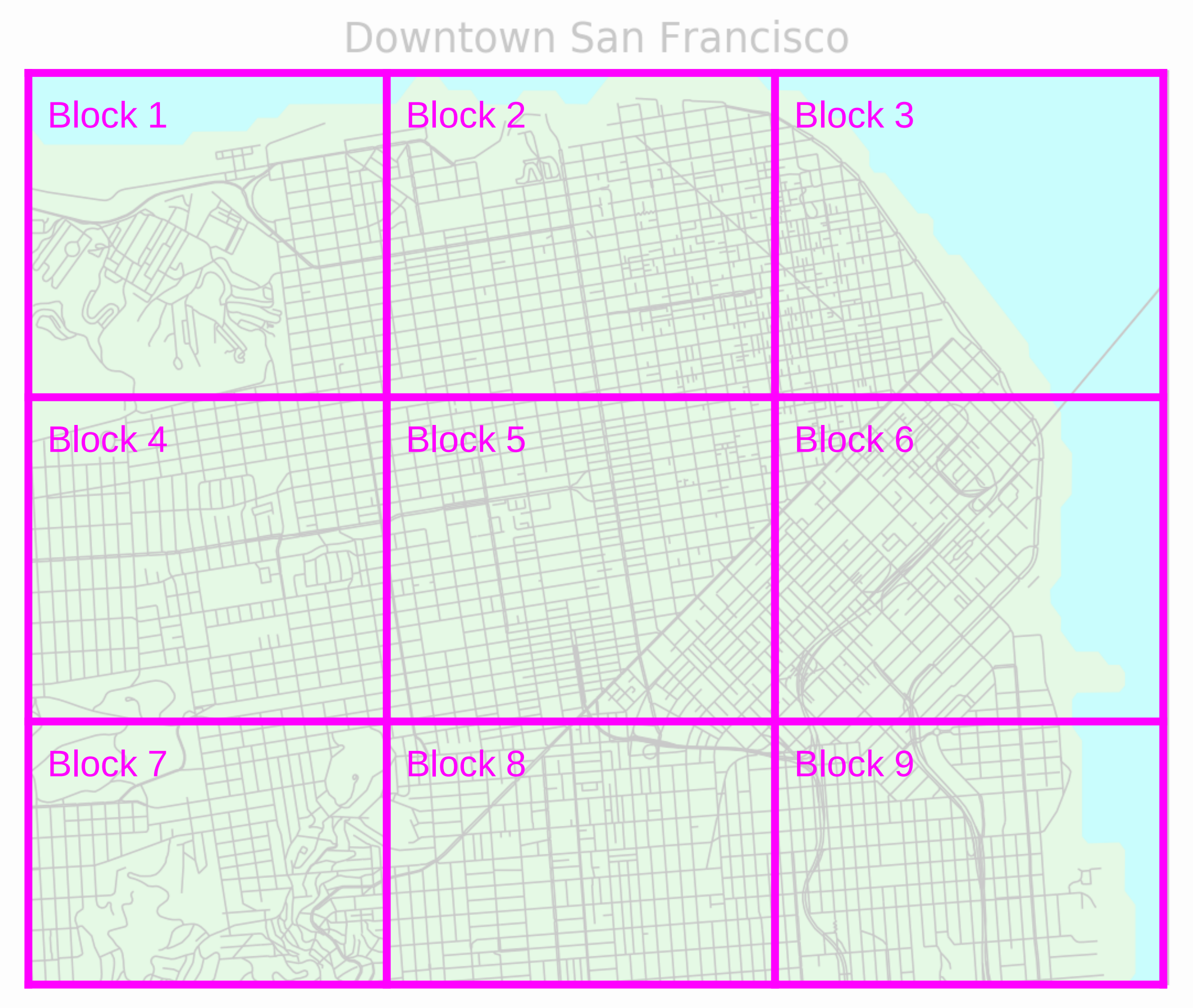
A pixel

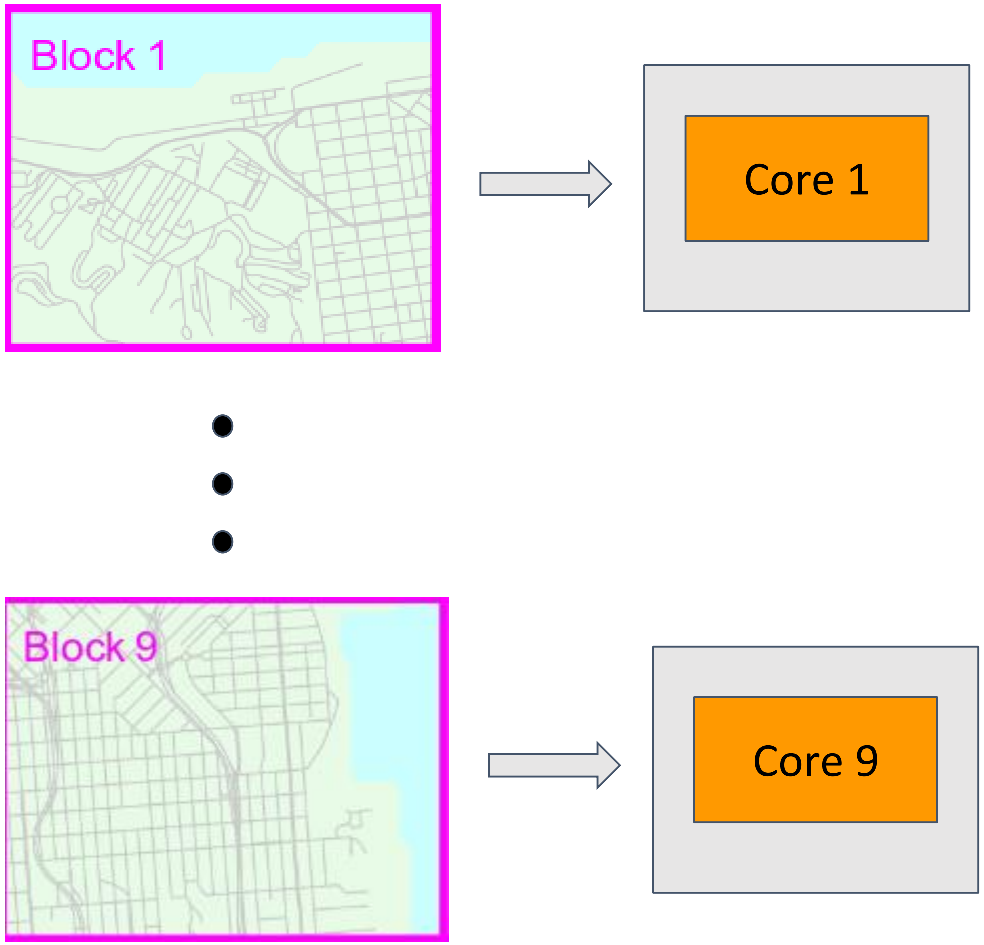
0	0	1	0
0	0	0	0
0	0	0	0
0	0	1	0
0	0	0	0

Grey scale image where each pixel represents the number of ride requests

Training Process

Every block is trained independently on a single core





Observation #2

Due to spatial independence of each block, all the blocks can be trained in parallel on many CPUs.

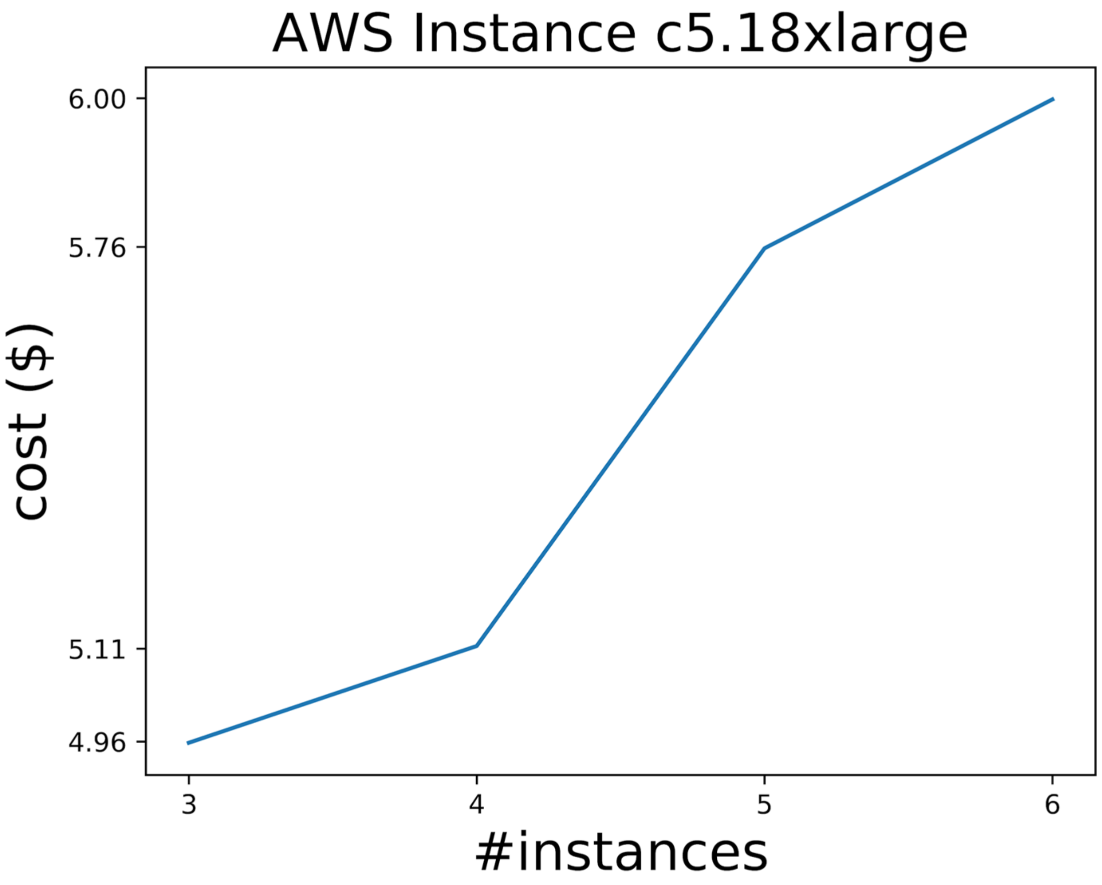
Computing Resources for Training on AWS

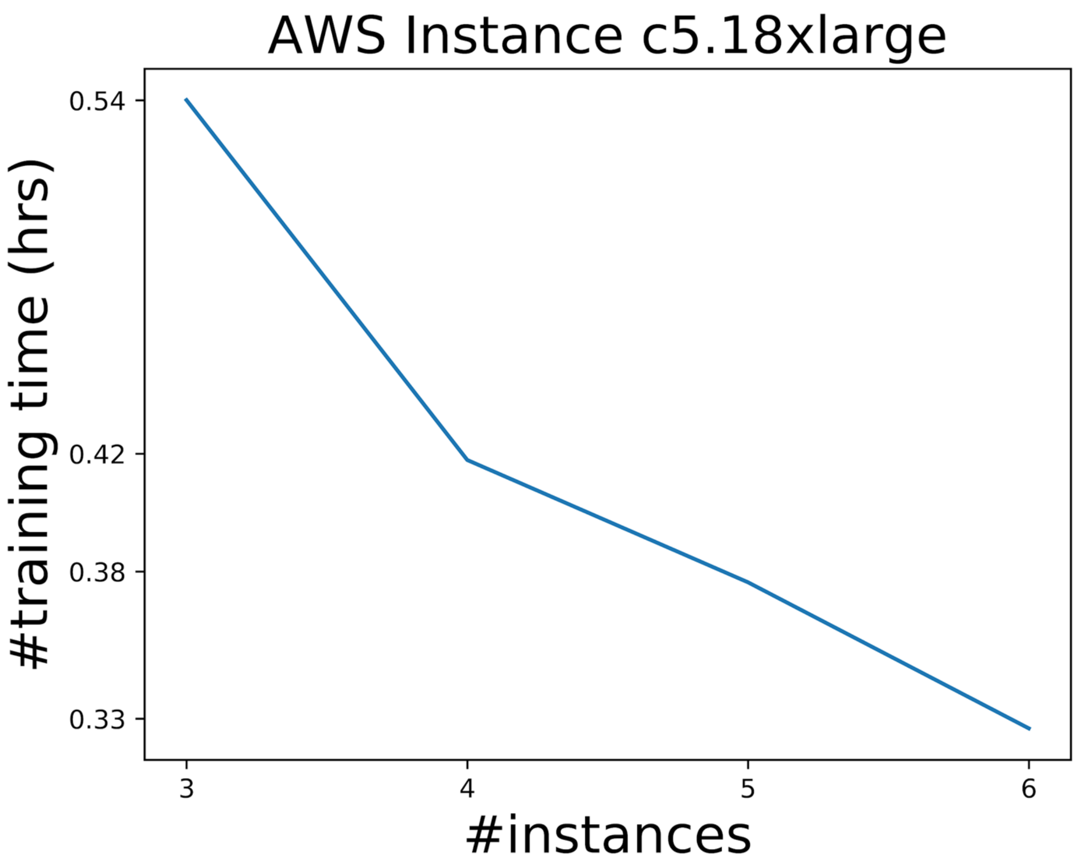
Experiments performed using --

1. c5.9xlarge - 36 cores; 3.0 GHz Intel Xeon Platinum 8000 Series

1. c5.18xlarge - 72 cores; 3.0 GHz Intel Xeon Platinum 8000 Series

Cost & Performance of Training GANs on AWS





Mobility21
A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING

Cost & Performance of Training GANs on AWS

Instance	Cost ($/hr)	Training Time (minutes)
c5.9xlarge x6	1.53	34
c5.18xlarge x6	3.06	19

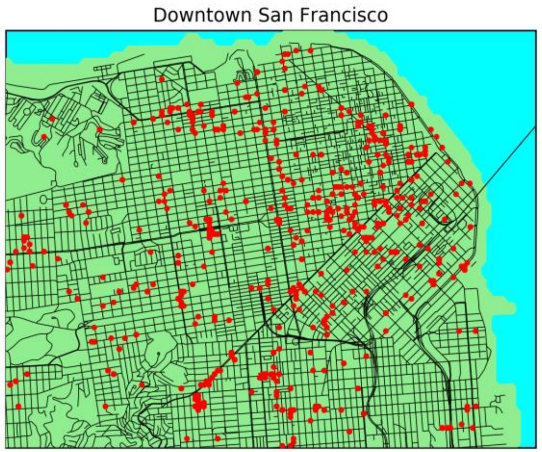
Outline

- Motivation
- Spatial and Temporal Variations
- Generative Adversarial Networks (GANs)
- Ride Requests to Images
- Experiments
- Conclusions

Training Workload for Different Cities

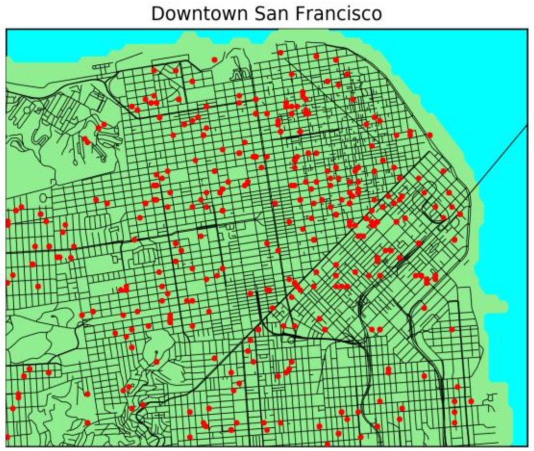
City	Number of blocks (geo divisions) for training
San Francisco	1402
New York	765
Chicago	1155
Los Angeles	1978

Results -- San Francisco Downtown



12am

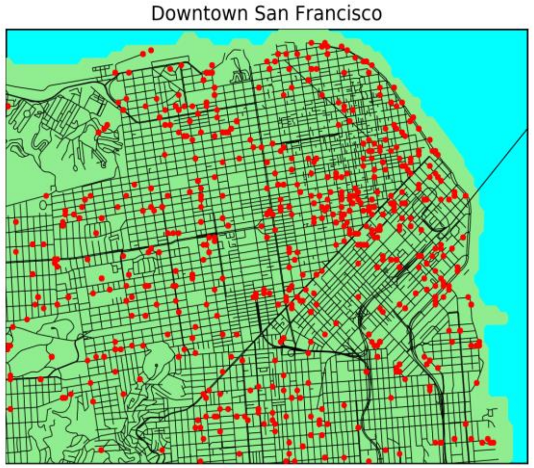
3am



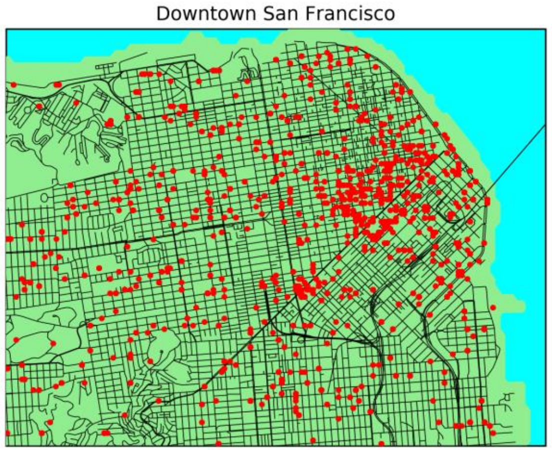
6am

9am

12pm



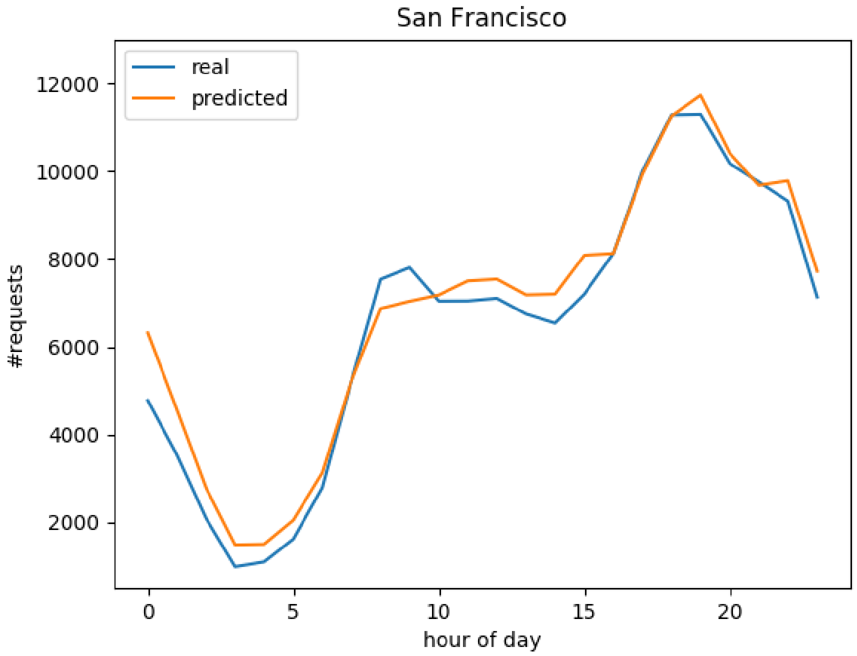
3pm



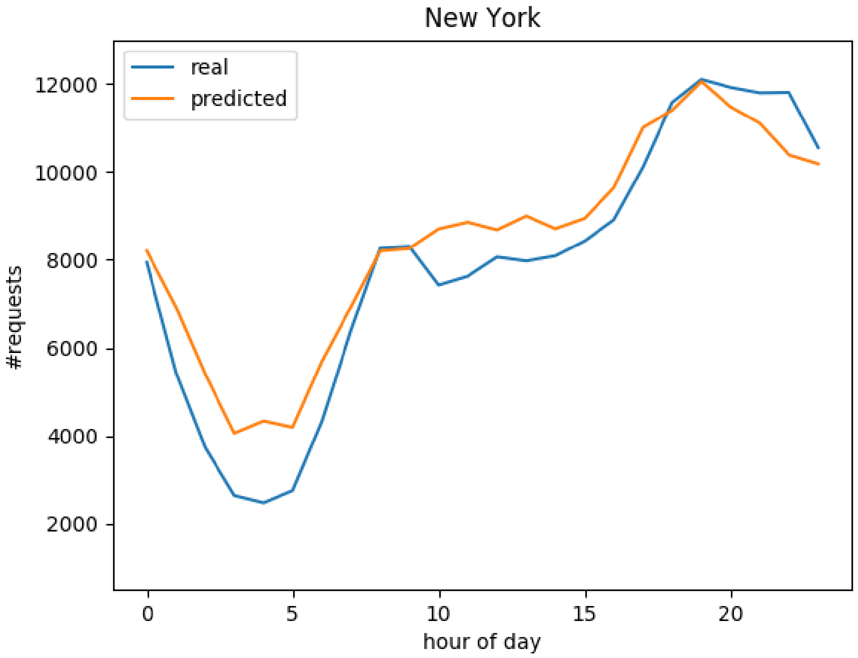
6pm

9pm

Temporal Validation: SF & NY



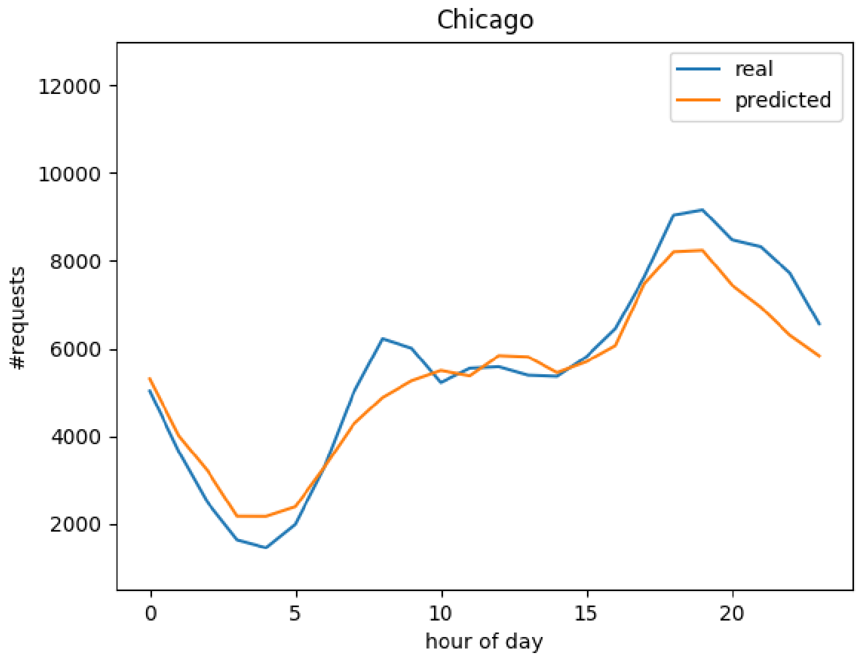
San Francisco



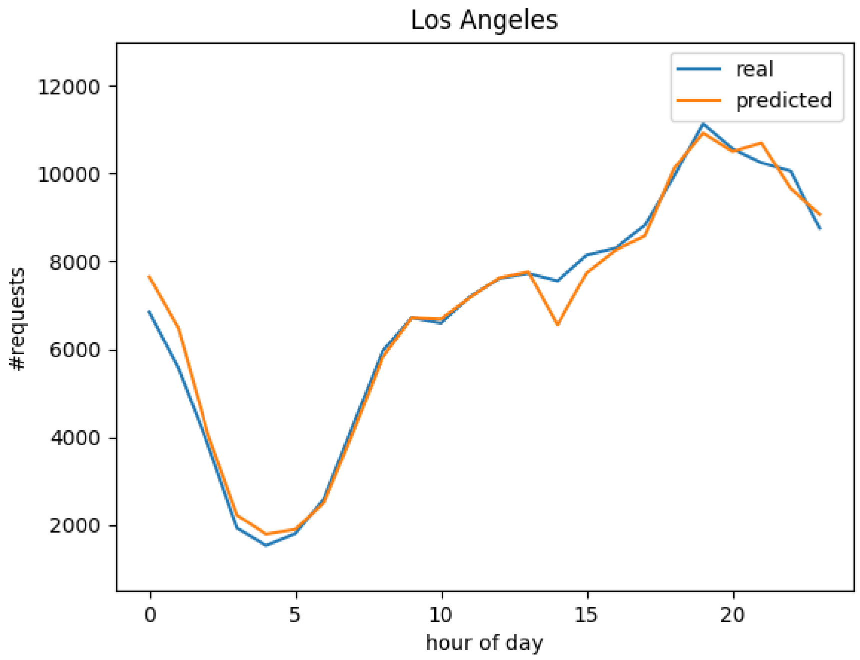
New York

Comparison of real and synthetic ride request volume for a day.

Temporal Validation: Chicago & LA



Chicago



Los Angeles

Comparison of real and synthetic ride request volume for a day.

Conclusions

- Highlighted a novel application of generating data for human mobility using GANs.
- Proposed model trains within **thirty minutes** for all four cities.
- Generated data sets match quite well the **spatial and temporal properties** of real data sets for all four cities.
- GANs generated data sets can be used by other researchers without privacy concern.

Questions

Mobility21

A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Electrical & Computer ENGINEERING