COVID-GAN+: Estimating Human Mobility Responses to COVID-19 through Spatio-temporal Generative Adversarial Networks with Enhanced Features

HAN BAO and XUN ZHOU, University of Iowa, United State
YIQUN XIE, University of Maryland, United State
YINGXUE ZHANG and YANHUA LI, Worcester Polytechnic Institute, United State

Estimating human mobility responses to the large-scale spreading of the COVID-19 pandemic is crucial, since its significance guides policymakers to give Non-pharmaceutical Interventions, such as closure or reopening of businesses. It is challenging to model due to complex social contexts and limited training data. Recently, we proposed a conditional generative adversarial network (COVID-GAN) to estimate human mobility response under a set of social and policy conditions integrated from multiple data sources. Although COVID-GAN achieves a good average estimation accuracy under real-world conditions, it produces higher errors in certain regions due to the presence of spatial heterogeneity and outliers. To address these issues, in this article, we extend our prior work by introducing a new spatio-temporal deep generative model, namely, COVID-GAN+. COVID-GAN+ deals with the spatial heterogeneity issue by introducing a new spatial feature layer that utilizes the local Moran statistic to model the spatial heterogeneity strength in the data. In addition, we redesign the training objective to learn the estimated mobility changes from historical average levels to mitigate the effects of spatial outliers. We perform comprehensive evaluations using urban mobility data derived from cell phone records and census data. Results show that COVID-GAN+ can better approximate real-world human mobility responses than prior methods, including COVID-GAN.

CCS Concepts: • Information systems → Geographic information systems; Spatial-temporal systems; • Computing methodologies → Machine learning;

Additional Key Words and Phrases: Mobility estimation, conditional generative adversarial networks, COVID-19

This article is funded in part by Safety Research using Simulation University Transportation Center (SAFER-SIM). SAFER-SIM is funded by a grant from the U.S. Department of Transportation's University Transportation Centers Program (Grant No. 69A3551747131). However, the U.S. Government assumes no liability for the contents or use thereof. Yiqun Xie is supported in part by NSF Grants No. 2105133 and No. 2126474, Google's AI for Social Good Impact Scholars program, and the Dean's Research Initiative Award at the University of Maryland. Yingxue Zhang and Yanhua Li were supported in part by NSF Grants No. IIS-1942680 (CAREER), No. CNS-1952085, No. CMMI-1831140, and No. DGE-2021871.

Authors' addresses: H. Bao and X. Zhou (corresponding author), University of Iowa, 14 MacLean Hall, Iowa City, IA 52242; emails: {han-bao, xun-zhou}@uiowa.edu; Y. Xie, University of Maryland, 1124 Lefrak Hall, 7251 Preinkert Dr., College Park, MD 20742; email: xie@umd.edu; Y. Zhang, Worcester Polytechnic Institute, 110 Institute Rd, Worcester, MA 01609; email: yzhang31@wpi.edu; Y. Li, Worcester Polytechnic Institute, Atwater Kent 130, 110 Institute Rd, Worcester, MA 01609; email: yli15@wpi.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

2157-6904/2022/01-ART27 \$15.00

https://doi.org/10.1145/3481617

27:2 H. Bao et al.

ACM Reference format:

Han Bao, Xun Zhou, Yiqun Xie, Yingxue Zhang, and Yanhua Li. 2022. COVID-GAN+: Estimating Human Mobility Responses to COVID-19 through Spatio-temporal Generative Adversarial Networks with Enhanced Features. *ACM Trans. Intell. Syst. Technol.* 13, 2, Article 27 (January 2022), 23 pages. https://doi.org/10.1145/3481617

1 INTRODUCTION

The COVID-19 pandemic has spread to over 200 countries globally, with 50.4 million confirmed cases and 1.26 million deaths as of November 8, 2020. The rapid spread of the infectious disease has posed grand challenges to policymakers due to the rising social conflicts between (1) the need to reduce new infections and protect public health and (2) the demand of reopening to avoid breakdown of economics and support essential needs in daily lives. As more policymakers start exploring mitigation of this dilemma with the staged reopening, a core mission is to avoid a major resurgence of infections caused by eased social distancing policies (e.g., stay-at-home orders, the limit of group gatherings, limit on restaurant capacity).

Currently, COVID-19-related policies are often informed by scientific projections of infection risks obtained from COVID-19 transmission dynamics models. While many parameters in infection dynamics models can be set using domain knowledge of COVID-19, a key parameter—*human mobility responses*¹—is often challenging to estimate due to complex and sometimes unknown social contexts, as well as limited training data under escalating COVID-19 conditions.

Responding to the urgent need by policymakers and public health experts, we make the first attempt to address the following *human mobility response estimation problem*: Given a set of inputs on contextual (e.g., population, **point-of-interest (POI)** counts), epidemic (e.g., COVID-19 cases), policy (e.g., stay-at-home orders) conditions. We aim at developing a machine learning model to estimate maps of human mobility responses by learning from existing ground truth data. Note, here the input conditions might not have been observed in the targeted area in the historical data.

Challenges. The human mobility response estimation problem problem has two major challenges. First, human mobility responses depend on many complex social-physical factors. While some factors are known and have related data available (e.g., population, POI distribution, policy), there exist many others (e.g., people's will or reaction over time, echo chamber effects, random events) that may be unknown, uncertain, random, or have no data available. In addition, the contribution of many known and unknown underlying factors to human mobility responses may vary from place to place due to cultural differences, economic environments, and so on. Moreover, the COVID-19 pandemic is a very recent and still-fast-changing event. The amount of related data is relatively small. This increases the difficulty of learning, especially when considering the random effects brought by the complex social context.

Second, based on observation from our previous work, we noticed that human mobility response data shows a skewed distribution with a long right tail. Values along the right tail are tremendous and account for a tiny share of the total number of training data. However, this small group of samples greatly affects generating performance due to the huge variances between ground truth and generated values. As a result, the relatively small data and randomness increase the difficulty of finding and learning to predict for the outliers.

¹In this work, human mobility responses are measured by the number of visits to point-of-interest (POI) such as grocery and hardware stores, restaurants, gas stations, and so on.

Prior Work. Due to the urgent need for effective control over the COVID-19 pandemic, there have been many works studying the importance of human mobility responses in COVID-19 transmission dynamics models as well as the effectiveness of early-phase social distancing measures (e.g., isolation, stay-at-home order, lock-down) on containing the spread of the disease [5, 10, 21]. There have also been studies evaluating the feasibility of scalable contact tracing methods in COVID-19 monitoring [8]. While these studies have demonstrated the importance of human mobility responses, they do not address the main challenges in mobility estimation or simulation (e.g., effects of unknown, uncertain, and random factors; limited training data). In terms of learning approaches, most regression-based estimators (e.g., linear and non-linear regression, kernelized support-vector regressor) rely on a fixed set of features and cannot consider unknown and uncertain factors. Furthermore, deep learning-based regression approaches (e.g., CNN) typically require a huge number of training samples, which are not available in the case of COVID-19.

In our recent work, we proposed COVID-GAN [4], a deep conditional **Generative Adversarial Network (GAN)** [10, 12], to estimate mobility under various real-world conditions such as COVID-19 severity and local policy interventions. COVID-GAN is effective in estimation human mobility patterns with a lower average error compared with other baselines. However, it does not consider the effects of spatial heterogeneity over the study area, therefore producing high error in certain regions. Also, COVID-GAN suffer from low accuracy when handling outlier locations with extremely high mobility values. These outliers indeed have high significance due to their high population density and vulnerability to the spread of COVID-19.

Contributions. This article is a significant extension of our recent work [4]. In this article, we address the above limitations of COVID-GAN by proposing an improved spatio-temporal conditional Generative Adversarial Network model with enhanced spatial features to take consideration of spatial heterogeneity when estimating human mobility response patterns to COIVD-19. We call the new Conditional Generative Adversarial Network model COVID-GAN+. The use of a conditional GAN [11] allows consideration of unknown and uncertain factors (i.e., modeled as latent factors). Moreover, we adopt a new training objective to learn the estimated mobility changes from historical average levels to mitigate the effects of local spatial hotspots.

The **new contributions** of this article are as follows:

- We measure the spatial heterogeneity of human mobility response by leveraging a new spatial feature layer using local Moran statistics [3]. This feature can help to better capture the spatial pattern and to improve the model capacity in generating fine-scale values.
- We reform the learning objective to learn the estimated mobility changes from historical average levels by introducing domain-knowledge forced conditions to mitigate the effects of spatial outliers. This correction mechanism can reduce the difficulty of learning outliers and improve solution quality.
- We perform comprehensive experiments on real-world COVID data to validate the solution quality improvements achieved by the proposed approach under different scenarios, and study the effects of related features on the quality of results.

Experiments using urban mobility data derived from cell phone records and census data show that COVID-GAN+ can achieve high accuracy when being tested against historical ground truth data, and can be used to estimate mobility trends under various conditions specified by users.

The rest of the article is organized as follows. Section 2 defines the problem. Section 3 details the methodology and model structure. We present evaluation results in Section 5 and further discuss and highlight the technical insights of the proposed solution framework in Section 6. Related works are summarized in Section 7, and the article is concluded in Section 8.

27:4 H. Bao et al.

2 OVERVIEW

Mobility estimation in the COVID-19 pandemic is a complex task requiring a variety types of information such as population, COVID-19 situation, and so on. Fortunately, many organizations have opened their data for research to help combat the pandemic. For example, SafeGraph [2], a leading provider of place and POI data, has started a COVID-19 Data Consortium to provide free access to their commercial data. Similarly, the **Center for Disease Control and Prevention (CDC)** also provides dynamic updates of COVID-19-related public heath statistics [1].

In the following part, we first introduce a set of basic concepts about our data modeling and then provide a formal problem definition.

2.1 Basic Concepts

Definition 2.1 (Spatial Grid Model $GRID_S$). A grid-discretization of a spatial field S, where each grid S represents a $l \times l$ square area. Given S, the location of any POI can be mapped into a grid cell. For example, the geographic region of interest in an estimation task.

Definition 2.2 (Contextual Conditions C_c). Relatively long-term attributes that are not necessarily specific to COVID-19, including population, median income, timestamp, and number of POIs (e.g., grocery stores, schools, gas stations, restaurants).

Definition 2.3 (Epidemic Conditions C_e). COVID-19 statistics from COVID-19 situation reports published authoritative public health organizations (e.g., CDC), including number of confirmed cases, COVID-19-related deaths, and so on.

Definition 2.4 (Policy Conditions C_p). Social distancing orders declared by officials during COVID-19 pandemic. Policies may play a key role in mobility changes. For example, strict stay-athome or shelter-in-place orders typically lead to a significant decrease in mobility.

Definition 2.5 (Human Mobility Responses M). In this work, **M** is measured by the number of visits to POIs (e.g., grocery stores, hardware stores, restaurants, gas stations), which according to our public health colleagues, is a major factor in COVID-19 transmission, since individuals tend to have closer distances and interact with each other at POIs.² Human mobility response is target of the estimation task.

Definition 2.6 (Generator G). A learned process used to generate a map of human mobility responses \mathbf{M}_G given a set of conditions.

Definition 2.7 (Discriminator D). Outputs a probability p_{real} that a map of human mobility responses is from real-world rather than a generator **G**.

2.2 Problem Statement

In this problem, we aim to learn a model to estimate the human mobility responses in each grid S during each timestamp. The timestamp could be an hour, a day, or a week. In this article, we select a week as the length of t. Time-variant features such as epidemic and policy conditions of week i are not visible for week i. So, we feed features from week i - 1 to estimate the human mobility responses on week i.

Inputs:

-A spatial grid model **GRID**_S for study area S, where $S = \{s_1, s_2, \dots, s_n\}$ is a spatial grid;

²In comparison, overlapping trajectories of private vehicles or passengers on private vehicles may raise less concerns in the spread process.

- -Contextual conditions \mathbf{C}_c (e.g., population, median income, POI counts, timestamp);
- -Epidemic conditions C_e (COVID-19 statistics);
- -Policy conditions C_p (e.g., social distancing policies);
- -Human mobility responses M.

Output:

-Generated maps (on $GRID_S$) of human mobility responses M_G under C_c , C_e , and C_D .

Objective:

—Minimizing estimation error on M_G .

Constraints:

- -Temporal resolution of estimation is aggregated at week-level;
- -Mobility estimation at week i can only use COVID-19 statistics achieved before week i;
- -Data is aggregated and does not contain private information.

In the scope of the present study, human mobility responses are measured by number of visits to POIs, and other related measures are currently out of scope and will be evaluated in future work.

In addition, the current temporal resolution of our analysis is set to week-level due to the very high variance/noise of mobility from day to day within a week. Thus, to put more confidence on estimation results, the scope in this article is to perform mobility estimation at an aggregated week-level as defined by the first constraint. The second constraint is added because changes in COVID-19 statistics (e.g., cases, death) can be considered as a direct or indirect result of mobility changes. Thus, in this problem, we make such post-information not accessible to mobility estimation for the same timestamp. However, this does not limit the use of COVID-19 statistics from previous timestamps, which may truly affect people's mobility responses.

3 BASELINE: COVID-GAN

In this section, we present the COVID-GAN model from our recent work [4] and use COVID-GAN as the building block for the new method proposed in this article. The COVID-GAN approach is inspired by the **conditional Generative Adversarial Network (cGAN)** [11]. cGAN is a natural modeling structure for human mobility response estimation in this COVID-19 scenario, because there exist a number of known and unknown/random underlying factors. With the cGAN structure, the known factors can be fed in as learning conditions and unknown/random variables can be represented by vectors of latent code. Our previous work makes the first attempt to formulate the COVID-19 human mobility estimation problem as a deep learning-based data generation problem.

3.1 Data Integration and Preprocessing

To construct the list of conditions for our input, for each grid cell in each week, we preprocess data collected from different sources with different geographic units.

Figure 1 shows a summary of the multi-view data we gathered from multiple sources based on definitions in Section 2.1 (i.e., contextual, epidemic, and policy conditions; human mobility responses). As we can see, most of the data is associated with different geographic units (e.g., census block groups, counties) due to different data sources or privacy protection concerns. Thus, further spatio-temporal data processing are needed before training and estimation.

First, to integrate all the data of various types and geographic units into the same format, and feed them into the conditional COVID-GAN, we adopt a commonly used space-partitioning approach [20, 40] to segment the input spatial domain into grid cells of size $1 \text{ km} \times 1 \text{ km}$, and all input data in Figure 1 are then segmented based on these grid cells. Note that some features are

27:6 H. Bao et al.

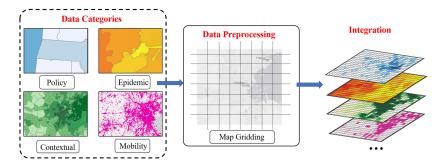


Fig. 1. Multi-source data and preprocessing for COVID-GAN.

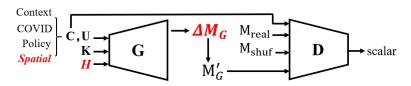


Fig. 2. COVID-GAN and COVID-GAN+.

re-scaled during this process. For example, population and median household income data are collected at the census tract level, and we linearly re-scaled the data using the corresponding area ratios between the area of the original census tract polygon and the proposed 10×10 grid cells. Similarly, COVID-19 statistics and policy data are collected at the county level. We assign each grid cell with the corresponding data on which county it belongs.

Since the temporal resolution of mobility estimation is week-level as discussed in problem definition (Section 2), we further merge all the daily information (e.g., mobility, number of cases and death related to COVID-19) into weekly aggregations for all grid cells.

Finally, we select spatial unit windows consisting of $s \times s$ grid cells with their corresponding features (i.e., conditions and constraints for COVID-GAN) to train our COVID-GAN.

3.2 COVID-GAN Architecture

This section presents our model structure to address the human mobility response estimation problem. Figure 2 shows the adversarial structure of the proposed COVID-GAN. COVID-GAN is composed of a conditional generator G and a discriminator D. To help improve the estimation quality, we propose a phase in the generator to constrain the estimated values using domain knowledge. Next, we will discuss the three key components in our COVID-GAN, i.e., the conditional generator, discriminator and domain-knowledge constraint. Note, Figure 2 illustrates the structures of COVID-GAN+, differences made by COVID-GAN+ based on COVID-GAN is highlighted in red.

3.2.1 Generator Phase-1: Conditions and Latent Factors. As shown in Figure 2, the input of conditional generator (skipping the dimension of "batch" for simplicity of description) includes a condition tensor $\mathbf{C} \in \mathbb{R}^{s \times s \times c_0}$, latent code tensor $\mathbf{U} \in \mathbb{R}^{s \times s \times u_0}$, and a domain knowledge tensor $\mathbf{K} \in \mathbb{R}^{s \times s \times k}$, where s is the side length of each spatial unit window introduced in Section 3.1, c_0 is the number of conditions (including contextual features, COVID statistics, and policy conditions), u_0 is the dimension of the latent code vector used to model the randomness in human mobility responses and k is the number of layers in domain constraint.

In generator, the **local-feature-projection (LFP)** layer is used to map existing features in the current layer L_i to those of the next layer L_{i+1} . The local-feature-projection layer is a special case of a convolutional layer in which the kernel size is $1 \times 1 \times F_i \times F_{i+1}$, where F_i and F_{i+1} are the number of features/channels in layer L_i and L_{i+1} , respectively.³ We use local feature projection layers instead of convolutional layers with larger kernel size mainly to reduce the boundary effects on grid cells of each spatial window. This is important in mobility estimations, since the final mobility map is typically an integration of many overlapping spatial windows (Section 4.4), which may lead to accumulated boundary effects and thus reduce generation quality.

The detailed architecture of the conditional generator is shown in Figure 6. Note, Figure 6 illustrates the model architecture of COVID-GAN and COVID-GAN+, differences made by COVID-GAN+ based on COVID-GAN is highlighted in red. In generator, the input tensors \mathbf{C} and \mathbf{U} are concatenated together and pass four local-feature-projection layers. The first three local-feature-projection layers are activated by **Rectified Linear Unit (ReLU)** and batch normalized. The last local-feature-projection layer is activated by hyperbolic tangent function, and we get an $s \times s$ matrix \mathbf{M}_G , where each entry represents the estimated human mobility of the corresponding grid cell. Then, domain-knowledge tensor \mathbf{K} are enforced on \mathbf{M}_G before the mobility values are fed into the discriminator to correct generated values and improve mobility estimation quality.

3.2.2 Generator Phase-2: Domain Knowledge Constraint. Recently, domain knowledge assisted learning has become an emerging trend, which can help reduce the difficulty of learning and overfitting with a small amount of training data and shows encouraging improvements in learning results. The domain knowledge ranges from observations (e.g., lake temperature versus water depth [18]) to social and physical theories (e.g., routine activity theory [9], engine combustion models [24]).

In COVID-GAN, we add a $s \times s \times k$ tensor K to represent domain knowledge that can be potentially used to correct or improve mobility estimations in phase-1. A domain constraint should satisfy the following requirements: (1) it contains finite layers (e.g., in our study, the layers of the domain constraint K corresponds to the last dimension of tensor K); (2) it is related to the input conditions C so that the usability of COVID-GAN does not depend on the constraint; and (3) for calculation simplicity, the constraint should be enforced or applied with a low-cost operation (e.g., a simple element-wise tensor calculation) using M_G and K. The third constraint is especially important in this context of deep learning, since the constraint enforcement operation will repeat in every iteration. In this study, we construct a constraint K with k = 1 satisfying all the three requirements mentioned above using a combination of **time, POI, and policy (TPP)** constraint.

Denote m_{xy} as mobility values at cell (x, y) in \mathbf{M}_G . We have $m_{xy} \in (-1, 1)$ as it is an output of $\tanh(\cdot)$ in the output layer. Since in COVID-GAN, human mobility at a grid cell is represented by the number of visits to POIs in a certain time slot (e.g., one week), m_{xy} equals to -1 when there is no visit in a cell, $m_{xy} > -1$ if (1) there exists POIs in the grid cell during the current time slot and (2) at least one of the POIs is allowed to open. To construct the corresponding TPP constraint \mathbf{K} with k = 1, we treat \mathbf{K} as an $s \times s$ matrix where each entry value equals to 1 if conditions (1) and (2) mentioned above are both satisfied; otherwise, the entry value is -1. Figure 3 shows an illustrative example of our TPP constraint.

The constraint can be then enforced by

$$\mathbf{M}_{G}' = \min(\mathbf{M}_{G}, \mathbf{K}). \tag{1}$$

 $^{^3}$ This type of layer has also been used in other networks (e.g., YOLO [28, 34]) for feature construction.

27:8 H. Bao et al.

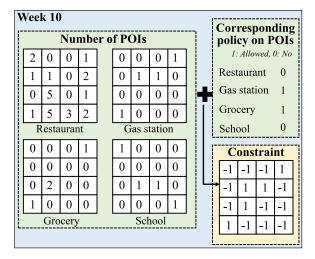


Fig. 3. An illustrative example of the TPP constraint.

By applying Equation (1), the mobility values in \mathbf{M}_G remain the same in if its corresponding TTP value is 1, or the values in \mathbf{M}_G are corrected to -1. Thus, this constraint overall only requires one data layer in \mathbf{K} and can be easily enforced by a single $\min(\cdot)$ function. This type of constraint corrections can potentially help the generator focus on valid cells and improve the generation quality.

3.2.3 *Discriminator.* Figure 6 shows the detailed architecture of the discriminator. Its input is a tensor of size $s \times s \times (c_0 + 1)$, where c_0 is the number of conditions (same as generator).

As shown in Figure 6, the input tensor can be created in three different ways:

- -Conditions concatenated with generated mobility M'_G ;
- -Conditions concatenated with their corresponding real mobility M_{real} ;
- —Conditions concatenated with mismatched real mobility M_{shuffle} .

Among the three kinds of inputs, only the second one is labeled with "real," whereas the other two are marked with "fake." The input tensor is then fed into the discriminator, and the goal of the discriminator is to learn to tell if the input is "real" or "fake."

The discriminator also contains four local-feature-projection layers, the output of the last local-feature-projection layer is reshaped into a vector and passes a fully connected layer as well as a Sigmoid function and outputs a final scalar, which indicates the probability whether the input of discriminator is "real" data.

3.3 COVID-GAN Training

The training of COVID-GAN is performed via adversarial confrontation between the generator and discriminator. The objective function of COVID-GAN is then formulated as an antagonizing bi-level min-max optimization problem with binary-cross-entropy:

$$\min_{G} \max_{D} f(G, D) = \mathbb{E}_{\mathbf{M} \sim P_{data}} [\log D(\mathbf{M}, \mathbf{C})] + \mathbb{E}_{\mathbf{U} \sim P_{U}} [\log(1 - D(G(\mathbf{C}, \mathbf{U}, \mathbf{K}), \mathbf{M}))].$$
(2)

Algorithm 1 presents the training process of COVID-GAN. The training of the discriminator uses the three types of (M,C) combinations as illustrated in Section 3.2.3: (M'_G,C) , $(M_{\rm real},C)$, and

($\mathbf{M}_{\text{shuffle}}$, \mathbf{C}). Denote η_D as the learning rate of discriminator, $\boldsymbol{\theta}_D$ as the parameters of discriminator, the loss function and the update rule of D are shown in Equations (3) and (4), respectively,

$$f_D = -\frac{1}{m} \sum_{i=1}^{m} \left(\log(1 - D((\mathbf{M}_G')^i, \mathbf{C}^i)) + \log(D(\mathbf{M}_{\text{real}}^i, \mathbf{C}^i)) + \log(1 - D(\mathbf{M}_{\text{shuffle}}^i, \mathbf{C}^i)) \right),$$
(3)

$$\theta_D = \theta_D + \eta_D \nabla f_D(\theta_D), \tag{4}$$

where m is the total number of samples in a batch, and index i refers to the ith sample.

Similarly, denote η_G as the learning rate of the generator G, and θ_G as the parameters in G, we have the loss function and update rule of G as

$$f_{G} = \frac{1}{m} \sum_{i=1}^{m} \left(\log(1 - D((\mathbf{M}_{G}^{\prime})^{i}, \mathbf{C}^{i})) \right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(\log(1 - D(G(\mathbf{C}_{i}, \mathbf{U}_{i}, \mathbf{K}_{i}), \mathbf{C}^{i})) \right),$$

$$\theta_{G} = \theta_{G} + \eta_{G} \nabla f_{G}(\theta_{G}).$$
(5)

ALGORITHM 1: COVID-GAN Training

Require:

- List of conditions C
- List of domain constraints K
- List of real mobility M_{real}
- Number of epochs epoch
- 1: G = initG()
- 2: D = initD()
- 3: **for** e = 1 to epoch **do**
- 4: **for** batch in $\{C, K, M_{real}\}$ **do**
- 5: $\mathbf{M}'_{G} = G(\text{batch.C}, rand(P_{\mathbf{U}}), \text{batch.K})$
 - {# for discriminator:}
- 7: $batch_{mis} = mismatchShuffle(batch)$
- 8: Update D using Equations (3) and (4)
- 9: {# for generator:}
- 10: Update G using Equations (5) and (6)
- 11: end for
- 12: end for

6:

4 MOBILITY ESTIMATION THROUGH COVID-GAN+

The baseline method COVID-GAN cannot well estimate the mobility in some regions due to spatial heterogeneity in the data (e.g., urban versus suburban) and outliers with extremely high mobility counts (e.g., downtown areas with large numbers of POIs). To address these issues, we propose a new model, namely, COVID-GAN+, to improve the estimation accuracy.

4.1 Quantifying Spatial Heterogeneity

We introduce spatial heterogeneity measures as a new condition to capture the inherent spatial structure among grid cells. Each grid cell has a time series of human mobility response (POIs visit

27:10 H. Bao et al.

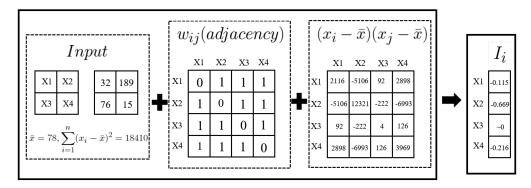


Fig. 4. An example of Local Moran's I calculation for 2×2 grids.

counts) through the study time period. We calculate the *Local Moran's I* [3] using human mobility response before COVID-19 outbreak to quantify the level of spatial heterogeneity between grid cell pairs in **GRID**_S. Local Moran's I measures the similarity between variables x_i and x_j , where x_i and x_j are human mobility responses value at grid cell i and j in our case. The Local Moran's I statistics \mathbf{I}_i is calculated by the following formula, where \overline{x} denotes the mean of x:

$$\mathbf{I}_{i} = \frac{\sum_{j=1}^{n} w_{ij} (x_{i} - \bar{x}) (x_{j} - \bar{x})}{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2}},$$
(7)

where w_{ij} are the elements in spatial weight matrix that determine the spatial relationship between location i and j, and is defined as $w_{ij} = 1$ if grid cell i and j are neighbors; otherwise, $w_{ij} = 0$. Figure 4 illustrates the calculation of Local Moran's I of a toy example with grid size 2×2 .

By utilizing Equation (7), I_i gives a statistic value for the association between grid cell i and all grid cells j. Unlike other similarity measurements, Local Moran's I is slightly different due to spatial calculation involvement. A negative value of I_i indicates that a grid cell has neighboring cells with dissimilar values; this grid cell might be an outlier. A positive value indicates that a grid cell has neighboring grid cells with similarly high or low human mobility responses; this grid cell is highly within a clustering. The benefit from this property is that applying Local Moran's I can help the model find outliers and capture clustering patterns in the study area.

Note that the Local Moran's I is a relative measure across neighborhood grid cells, instead of directly use the value, it is necessary to determine the statistical significance. We compute a Z-score using the following equation where $Var[I_i]$ is the variance of values:

$$\mathbf{z}_{Ii} = \frac{I_i - E[I_i]}{\sqrt{Var[I_i]}},\tag{8}$$

where

$$E[I_i] = -\frac{\sum_{j=1, j \neq i}^{n} w_{ij}}{n-1},$$
(9)

$$Var[I_i] = E[I_i^2] - E[I_i]^2.$$
 (10)

The above process will generate the Local Moran's I z-score for each grid cell, which is directly assigned to each grid cell as a new spatial correlation condition. We finally constructed eight features for COVID-GAN+, which are from four categories.

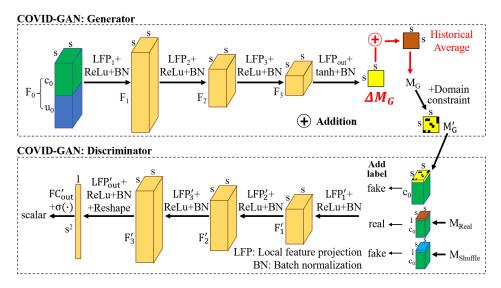


Fig. 5. COVID-GAN and COVID-GAN+ Network Architecture.

4.2 COVID-GAN+ Architecture

Rare events such as the COVID-19 pandemic typically have limited availability of training data but require timely policymaking and intervention. The "adversarial" setting in GAN mitigates this challenge by a self-improving discriminator. Using conditional GAN allows generating a variety of human mobility responses under the same input conditions of known factors.

As discussed above, our previous model does not handle spatial heterogeneity explicitly, although we incorporate a spatial correlation condition, model performance still greatly affected by varying environmental conditions. One difficulty is detecting and estimating outliers. These spatial outliers are very informative in our case, while they are also challenging to detect due to small samples. To help improve the model capacity in dealing with outliers, we reform the learning objective of the conditional generator G to learns the estimated mobility changes from historical average levels by forcing domain constraints to reduce the effects of outliers. Figure 2 shows the overall structure of COVID-GAN+. Next, we will discuss the generator and discriminator in COVID-GAN+.

4.2.1 COVID-GAN+ Generator. To tackle the challenge of spatial heterogeneity, we introduce the spatial correlation condition in constructing the input of a conditional generator. As Figure 2 shows, C contains four conditions (including contextual, features, COVID statistics, policy, and spatial heterogeneity conditions).

Figure 5 shows the overall architecture of COVID-GAN+. In generator, the input tensors ${\bf C}$ and ${\bf U}$ are concatenated together and pass four local-feature-projection layers. We get an $s \times s$ matrix M_G , where each entry represents the estimated human mobility of the corresponding grid cell. Between the output M_G and the domain-knowledge tensor ${\bf K}$, we implement an additional environmental constraint ${\bf H}$ to reduce the difficulty of generating mobility estimation in dense regions. This step forces the model learn the mobility changes, which denote as ΔM_G .

To accomplish this, we adapt the domain knowledge constraint structure from COVID-GAN in conducting the environmental constraint. We construct a $s \times s \times k$ tensor **H**, where k = 1 represents the environmental constraint that can be used to improve mobility estimation in the generator, particularly the quality in estimating outliers.

27:12 H. Bao et al.

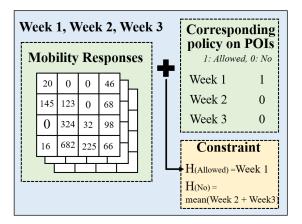


Fig. 6. An illustrative example of the environmental constraint.

Denote m_{xy} as mobility values at cell (x, y) in \mathbf{M}_G . We have $m_{xy} \in (-1, 1)$ as it is an output of $\tanh(\cdot)$ in the output layer. With the purpose of learning the ΔM_G , we treat \mathbf{H} as an $s \times s$ matrix with a value equal to the conditional historical average mobility response controlled by policy conditions.

We define the value of constraint **H** as follows:

$$\mathbf{H} = \bar{M} \times C_p + |C_p - 1| \times \bar{M},\tag{11}$$

where $C_p \in \{0, 1\}$, indicating if there is policy implemented. \bar{M} represents the historical average mobility responses. By applying Equation (11), the values of **H** are calculated by given policy condition in each grid cell. Then constraint **H** can be enforced by

$$\mathbf{M}'_{G} = \mathbf{M}_{G} + \mathbf{H}(\mathbf{H} \in (0,1)).$$
 (12)

By applying Equation (12), the mobility values in M_G are squeezed to a smaller range and represent the mobility response changes ΔM_G . The overall constraint only requires one data layer in **H** and can be easily implement through a pixel-wise AND operation.

This constraint correction can potentially help the generator focus on learning the mobility changes. Given conditional historical average values for grid cells under different policy conditions, the variance between common and outlier cells is reduced, improving the model's capacity to detect and estimate outliers. In generator, environmental constraint layer is followed by the domain knowledge constraint layer, then the \mathbf{M}_G' fed into a discriminator.

4.2.2 COVID-GAN+ Discriminator. Figure 5 shows the detailed architecture of the discriminator. Its input is a tensor of size $s \times s \times (c_0 + 1)$, where c_0 is the number of conditions (same as generator), and the one extra layer added to it contains mobility estimations.

We adapt the structure of COVID-GAN discriminator. The discriminator also contains four local-feature-projection layers. Within the discriminator, the hidden layers first construct local features to form signatures of "real" and "fake" distributions and then combine them into a single scalar value $\in (0,1)$ representing the probability of an input being generated by a real distribution.

4.3 COVID-GAN+ Training

The training of COVID-GAN+ is carried out by rotated training of them. We adapt the objective function of COVID-GAN and add the correction layer **H** in generator. The objective function is

ALGORITHM 2: COVID-GAN+ Training

```
Require:
     • List of conditions C

    List of domain constraints K

     • List of environmental constraints H

    List of real mobility M<sub>real</sub>

    Number of epochs epoch

 1: G = initG()
 2: D = initD()
 3: for e = 1 to epoch do
        for batch in {C, H, K, M<sub>real</sub>} do
            \Delta M_G = G(\text{batch.C}, rand(P_U), \text{batch.H}, \text{batch.K})
            \mathbf{M}_G = \Delta M_G + \mathbf{H}
           \mathbf{M}_G' = \min(\mathbf{M}_G, \mathbf{K})
 7:
            {# for discriminator:}
            batch_{mis} = mismatchShuffle(batch)
            Update D using Equations (14) and (15)
10:
            {# for generator:}
11:
12:
            Update G using Equations (16) and (17)
        end for
14: end for
```

formulated as follows:

$$\min_{G} \max_{D} f(G, D) = \mathbb{E}_{\mathbf{M} \sim P_{data}} [\log D(\mathbf{M}, \mathbf{C})]
+ \mathbb{E}_{\mathbf{U} \sim P_{U}} [\log(1 - D(G(\mathbf{C}, \mathbf{U}, \mathbf{H}, \mathbf{K}), \mathbf{M}))].$$
(13)

Algorithm 2 presents the training process of COVID-GAN. The training of the discriminator uses the three types of (M, C): (M'_G, C) , (M_{real}, C) , and $(M_{shuffle}, C)$. Denote η_D as the learning rate of discriminator, θ_D as the parameters of discriminator, the loss function and the update rule of D are shown in Equations (14) and (15), respectively,

$$f_D = -\frac{1}{m} \sum_{i=1}^{m} \left(\log(1 - D((\mathbf{M}_G')^i, \mathbf{C}^i)) + \log(D(\mathbf{M}_{\text{real}}^i, \mathbf{C}^i)) + \log(1 - D(\mathbf{M}_{\text{shuffle}}^i, \mathbf{C}^i)) \right),$$

$$\theta_D = \theta_D + \eta_D \nabla f_D(\theta_D),$$
(15)

where *m* is the total number of samples in a batch, and index *i* refers to the *i*th sample.

Similarly, denote η_G as the learning rate of the generator G, and θ_G as the parameters in G, we have the loss function and update rule of G as

$$f_G = \frac{1}{m} \sum_{i=1}^{m} \left(\log(1 - D((\mathbf{M}_G')^i, \mathbf{C}^i)) \right)$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(\log(1 - D(G(\mathbf{C}_i, \mathbf{U}_i, \mathbf{H}_i, \mathbf{K}_i), \mathbf{C}^i)) \right),$$
(16)

$$\theta_G = \theta_G + \eta_G \nabla f_G(\theta_G). \tag{17}$$

27:14 H. Bao et al.

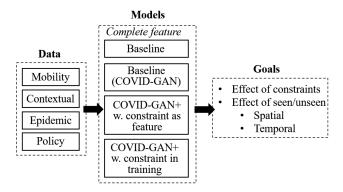


Fig. 7. Overall evaluation framework.

4.4 COVID-GAN+ Estimation

The goal of this step is to generate the mobility map of the entire study area (e.g., a city or county) from the estimations of $s \times s$ spatial unit windows. Since COVID-GAN+ would generate the mobility maps of spatial unit windows in different time slots and areas, we present two schemes to generate the final map of mobility estimation.

Single-draw-based sliding window: In this scheme, we move a $s \times s$ sliding window across the whole target area to prepare all conditions and constraints for the generator G. Then, using these well-prepared data, we train the COVID-GAN+ and generate a single $s \times s$ mobility estimation result for each unit window (i.e., a single draw from the distribution with latent factors), and we build the whole map with all the generated results for unit sliding windows. Since each grid cell in the target area may have multiple generated results due to overlaps among the sliding unit windows, the final mobility value in a grid cell is the average of the generated results.

Multiple-draw-based sliding window: This scheme takes w draws of the same window from the generator G instead of a single draw. The results of the multiple draws are averaged before being integrated into the estimation of the original study area.

In our experiments, we use the multiple-draw version to reduce random effects in comparison. The single-draw version is better suited for scenarios when a decision maker would like to explore potential variations in human mobility responses.

5 EVALUATION

The overall evaluation framework is shown in Figure 7. Our experiments aim to answer the following questions:

- —How does COVID-GAN+ performs compared to the COVID-GAN and other baseline models in terms of solution quality?
- —What are the effects of features and constraint layer in COVID-GAN+ on the performance?
- —What is the effect of training on temporally seen versus temporally unseen samples?
- —What is the effect of training on spatially seen versus spatially unseen samples?

The second question aims to assess the learning effectiveness of environmental constraint layer in COVID-GAN+ on the performance. The third question aims to compare the quality of generated mobility responses between spatial regions that are seen and unseen in the training data (either temporally seen or not). Similarly, the fourth question aims to evaluate the same effect along the spatial dimension.

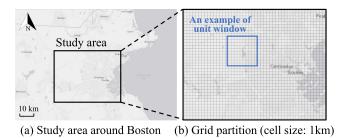


Fig. 8. Partitioning of Boston study area.

5.1 Boston Dataset

For human mobility responses estimation, we collect data from four data sources in 1 for Boston. The study area of Boston used in the experiments is shown in Figure 8, where Figure 8(a) visualizes the geographic context of the area around Boston, MA, and Figure 8(b) shows the grid partitioning of the space. The dimension of the study area is $37 \, \text{km} \times 48 \, \text{km}$ and each grid cell has a size of $1 \, \text{km} \times 1 \, \text{km}$. An example of a 10×10 unit window is illustrated by the blue box in Figure 8(b). The duration of data is from 03/02 to 05/24, covering 12 weeks in total.

The original POI dataset is aggregated daily and obtained by collecting the anonymous location from mobile devices. Each location includes the latitude and longitude of the corresponding device at a given point in time. Then, the location information is used to determine the visits to points of interest [2]. The study area has over 2 million population in total and 26,054 POIs. We further aggregate the daily mobility to weekly. During the 12 weeks, the total number of POI visits is over 6 million. For each grid cell in the study area, the value of mobility is the total number of POI visit counts in a week. A "closure-of-public-venues" policy was implemented starting from 03/23 and lasted till the end of the 12-week period.

5.2 Evaluation Metrics

We evaluate the performance of COVID-GAN+ by using the following measures: **mean absolute error (MAE)** and **rooted mean-square error (RMSE)**:

$$MAE = \frac{1}{n} \sum_{i=0}^{n} |y_i - \hat{y}_i|,$$
 (18)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y_i})^2},$$
(19)

where y_i is the ground truth and \hat{y}_i is the predicted values of mobility responses at each grid cell. To evaluate the model's performance more comprehensively, we use KL-divergence to indicate the similarity between the learned distribution and ground truth distribution by giving different bin sizes. For two distributions \mathbf{P} and $\hat{\mathbf{P}}$, the KL divergence is defined as follows:

$$\mathbf{D_{KL}}(P||\hat{P}) = \sum_{i=1}^{N} P(y_i) log\left(\frac{P(y_i)}{\hat{P}(\hat{y}_i)}\right),\tag{20}$$

where we obtain the two distributions by calculating \mathbf{M}_G' and \mathbf{M}_{real} histogram and applying different bin sizes in generating histograms to evaluate the matching between ground truth distribution and learned distribution. Besides, we record the total mobility estimated in the full study area and the sum of estimated mobility differences for all grid cells in the full study area.

27:16 H. Bao et al.

Model		HA	Smoothing	Ridge	MLP	cGAN	COVID-GAN	COVID-GAN+	Ground Truth
	RMSE	267	368	203	549	394	351	147	_
With conditional	MAE	61	169	139	256	157	167	65	_
HA as feature	Total	522,274	452,596	335,225	216,089	386,791	402,471	399,031	422,223
	Difference	100,051	30,373	-86,998	-206,134	$-35,\!432$	-19,752	-23,192	_

Table 1. Performance Comparison of Different Approaches

5.3 Baseline Methods

We compare our model with the following six baselines:

- Historical Average (HA): The average value of human mobility responses at the same time
 interval. In this article, we calculate the historical average using mobility from temporally
 seen weeks.
- Spatial smoothing with neighborhood regions [13]: For a given time slot, this method uses the mobility response values of the nine closest regions to compute a mean value as the resulting estimation. Note, we use available data in the training set to estimate for this method.
- **Regression** [16]: Ridge regression is applied to estimate the mobility responses in each grid cell given the condition list as predictors.
- cGAN [11]: A standard Conditional Generative Adversarial Network without any constraint layers. The generator and discriminator both pass through three fully connected layers, and output of the generator is activated by Tanh, and the output of the discriminator is fed to Sigmoid function.
- MLP: Multiple Layer Perception. We use three-layer MLP and select ReLU as the activation function.
- **COVID-GAN** [4]: COVID-GAN, a spatio-temporal Conditional Generative Adversarial Network including a domain-constraint correction layer in the generator to reduce the difficulty of learning proposed in our recent work.

5.4 Estimation Quality Evaluation

Here, we aim to answer the three questions summarized at the beginning of Section 5. Figures 11 and 12 show a comparison of results of different approaches in different scenarios. Specifically, the first column (i.e., (a1) and (a2)) in both figures shows the ground truth, and the second and third columns show the results of the proposed COVID-GAN+ as well as the baseline approach COVID-GAN without environmental constraint-based correction). In Figures 11 and 12, we highlight the results of the candidate methods for a sub-region in the study area (the same region is used for both temporally seen and unseen data, and spatially seen and unseen data, respectively). The results for the full Boston study area are shown in Table 1 and Figure 9. To comprehensively compare the performances between COVID-GAN+ and baseline methods, we reform the conditional historical average value used in the environmental constraint layer into a feature layer and fed into all the methods as a predictor.

The colors used in map symbologies are classified using quantiles extracted from the ground truth (i.e., 0th, 25th, 50th, 75th, and 100th), a typical approach for enhanced map visualization. To reduce random effects in the comparison, all results are based on the multiple-draw-based approach (10 repetitive runs) described in Section 4.4.

The performance of proposed COVID-GAN+ and baseline methods: Table 1 shows the statistical results of seven candidate approaches. Since COVID-GAN+ utilizes the environmental constraint layer in training, we fed conditional historical average value to all the baseline models as a predictor, and for COVID-GAN, we construct this information as a condition. The COVID-GAN+

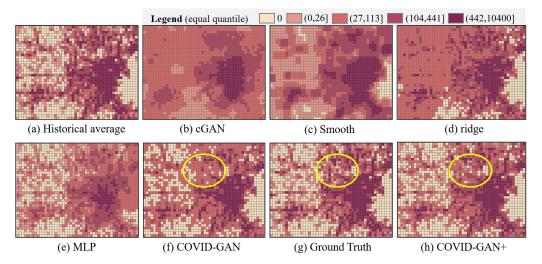


Fig. 9. Mobility estimation results of the full Boston study area (with conditional HA as a predictor in baseline methods).

achieves the best performance with the lowest RMSE, and the performance is greatly improved by comparing the RMSE and MAE with COVID-GAN. Historical average performs better than other baseline methods but not as good as our proposed approach. These two observations reveal that the historical average is a good method for mobility estimation with low error due to the periodicity pattern in human mobility patterns.

Table 1 also shows the total estimated mobility (i.e., the total number of POI visits) and differences between the ground truth and the models for the full study area computed using the ground truth, COVID-GAN+, and the baselines. The total mobility (number of visits) of the ground truth, COVID-GAN+, and baseline are 422,223, 399,031, and 402,471, respectively. The differences between the ground truth and the methods are -23,192 and -19,752 for COVID-GAN+ and COVID-GAN, respectively. We can see that the overall estimation of COVID-GAN and COVID-GAN+ are very close to that of the ground truth. The COVID-GAN+ outperforms in the MSE and RMSE; however, COVID-GAN+ loses in Total. The possible reason is that COVID-GAN generates mobility values with large variance compared to COVID-GAN+, increasing the total amount of mobility estimation. Generally, compared to other baselines, both COVID-GAN and COVID-GAN+ have much better performance. To evaluate the distribution-closeness between the mobility values estimated by the ground truth and the methods, we further compute the Kullback-Leibler divergence [22] and the results are shown in Figure 10. The x-axis is the number of equal-size bins used to discretize the mobility estimations needed for the computation. As we can see, COVID-GAN+ achieves lower KL-divergence values compared to the baselines consistently for different numbers of bins, and the historical average performs well in this measurement. From the statistical results, we notice the strong benefit of using the historical average for human mobility estimation.

Visualization: Besides statistical results, we compare the solution quality of the seven candidate approaches through map visualization. Figures 9(a) to 9(e) show the results of baseline methods, and Figures 9(f) and 9(g) show the COVID-GAN, ground truth, and results of proposed COVID-GAN+ for the full Boston study area for the final week in the data (not used in training). The yellow circle highlights an example of the differences. Here COVID-GAN+ generates fine-scale mobility values that are closer to the ground truth. As we can see, the mobility pattern generated by the COVID-GAN+ can capture the details in the distribution of human mobility responses better than

27:18 H. Bao et al.

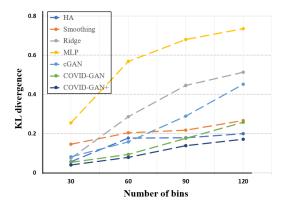


Fig. 10. Kullback-Leibler divergence (with conditional HA in baseline methods).

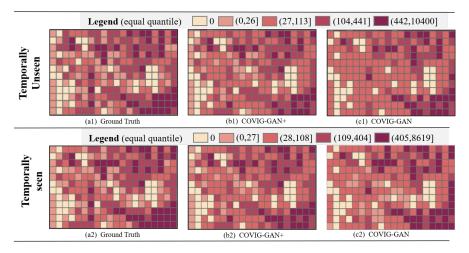


Fig. 11. Detailed mobility estimation results in a sub-region of the study area for both temporally seen and unseen data (with conditional HA).

other baseline methods. The reason may be that COVID-19-related policy has a discouraging effect on people's will to get out. Forcing the model to learn based on the conditional historical average value controlled by policy leads to stronger suppressing force on the value and an increase in the estimation. Note that MLP and standard cGAN generate mobility response values in the sea region locates on the middle of the right side. This supports the usage of the domain-knowledge constraint correction mechanism in COVID-GAN+ models.

COVID-GAN+ versus COVID-GAN: As we can see, in Table 1 and Figure 9, results of COVID-GAN+ in Figure 9(h) can better approximate the mobility distributions in the ground truth compared to the results of the COVID-GAN method in Figure 9(f), where the pattern of a large number of cells around the downtown dense region are well depicted. Besides, the trend remains the same for both temporally seen and unseen data in Figure 11 and spatially seen and unseen data in Figure 12. From Table 1 we can see the RMSE of COVID-GAN+ improves 62%, and there is a 63% improvement in MAE. As we discussed, outliers are composed of a small number of samples; however, they largely affect the estimate accuracy. By introducing the environmental constraint layer, the difficulty of learning outliers is reduced and, at the same time, there is an improvement in the overall estimation quality.

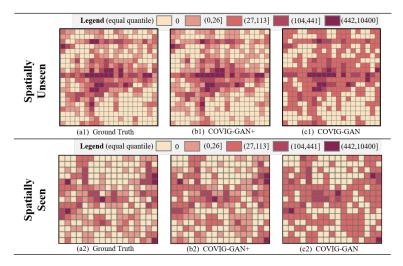


Fig. 12. Mobility estimation results in spatially unseen (first row) and seen regions (second row) of the study area (with conditional HA).

Temporally seen versus temporally unseen: In this comparison, "temporally seen" refers to the time periods (i.e., weeks) in the training data and have been seen by COVID-GAN+. In contrast, temporally unseen refers to data of weeks that are outside the training data.

We expect that our model can generalize reliable human mobility responses estimation given different condition combinations. The purpose of temporally unseen versus temporally seen is to examine the model estimation accuracy given new temporally unseen conditions. To make the test more realistic, the timestamp of unseen data must also be strictly after all timestamps in the training data so that the model does not try to estimate the past based on the future. We always feed prior conditions to estimate the following week's mobility. We use week 12 data as a testing week to evaluate the model performance with unseen conditions temporally.

In this experiment, we use the first 11 weeks of data to train the candidate models and leave the 12th week out as "temporally unseen" data. According to the results in Figure 11, we can see that COVID-GAN can maintain a good description of the overall pattern. The grid cells on the center of the sub-region in Figure 11 in COVID-GAN+ model have a higher solution quality while COVID-GAN gives a blur description of this sub-region. This is especially important in assisting policymaking during the ongoing pandemic when model being used to estimate human mobility responses for the unseen week.

Spatially seen versus spatially unseen: To generate spatially unseen regions, we crop a $20 \, \text{km} \times 20 \, \text{km}$ sub-region (i.e., a 20×20 sub-grid) off the total geographic space shown in Figure 8, and condition combinations in this sub-region are not seen by COVID-GAN+ during entire training phase. This eliminates about one-third of the total amount of training samples (overlaps with the sub-grid are not allowed). Figure 12 shows the comparison of results by COVID-GAN+ and the COVID-GAN. Comparing results in the two rows, COVID-GAN+ achieves a better estimation for the first row when the data is spatially seen, and the results of the COVID-GAN cannot well approximate the details in the mobility distribution.

6 DISCUSSION

In this section, we discuss a few data and technical aspects of COVID-GAN+ in a more holistic manner.

27:20 H. Bao et al.

Generation versus prediction: As a conditional generative neural network, COVID-GAN+ is more of a generator than predictor by definition, since the generation process involves randomly sampled latent factors to model unknown or uncertain factors in human mobility responses. In other words, the generated mobility results may not necessarily remain the same under the same condition (results in our experiments are averaged from multiple simulations). This allows policy-makers or epidemiologists to evaluate different possibilities of human mobility responses during decision-making (e.g., change of policy). In contrast, a prediction model typically produces the same result for the same data once trained, and it is better suited when all or most of factors of human mobility responses are well-understood, stable and available in data.

Spatial variability: While our experiment results show that COVID-GAN+ performed reasonably well for a spatially unseen region, that region is relatively small and is still adjacent to the spatially seen data. Thus, more research is needed to examine COVID-GAN+'s performance for regions that are more distant to each other (e.g., different states, different countries). Given the typical social and physical differences in different geographic areas, new approaches may be needed to explicitly handle the spatial variability at larger geographic scales [15].

Data quality: In this work, we did not investigate the effect of data quality (e.g., missing data, inaccuracy, anomaly) of input conditions, which may have bigger impacts in rural areas where the total volume of data tend to be smaller. This needs to be further investigated in future research to improve the generality of COVID-GAN+.

7 RELATED WORK

COVID-19. There have been many studies [5, 7, 10, 21] exploring the interplay between human mobility responses, social distancing policies, and transmission dynamics in response to the COVID-19 pandemic. For example, it was shown by Reference [21] that strict implementation of social distancing policies can reduce mobility and substantially mitigate the spread of COVID-19. A U.S. mobility change map was created in Reference [10] to increase risk awareness of the public and visualize dynamic changes in mobility as COVID-19 situation and policy evolves. Reference [7] measures the effectiveness of non-pharmaceutical interventions introduced by governments across Europe using the changes of mobility. Studies [8] have also explored the feasibility of utilizing contact tracing to control the spread of the disease through simulated synthetic data and real-world smartphone trajectories. These studies are timely in showing the important role played by mobility in the spread of COVID-19, but they do not address the challenges in real-world mobility estimation/simulation(e.g., effects of unknown, uncertain, and random factors), and they analyze the mobility changes in city or country scale. Besides, these studies have not explored the potential use of deep learning-based generative models to assist the estimation.

Deep Learning for Spatio-temporal Prediction. There have been many deep learning-based techniques developed for spatio-temporal data. For example, LSTM were widely used in traffic accident prediction [37], and flow prediction [26, 38], due to its capability in capturing spatio-temporal correlation and thus provide good prediction results. Geospatial object mapping [34–36], taxi driver behavior imitation [39], taxi demand [12], travel-time estimation [33], dispersal event forecasting [31], and so on, all combine the deep learning model with spatio-temporal perspective in their model design and obtain good performance. Most of these spatio-temporal deep learning techniques typically are stationary predictors (i.e., same result from two runs on same data) rather than generative models, and they do not consider the unknown factors in prediction, and their performance relies on large data sets. Besides, they do not leverage domain knowledge-based constraints to assist learning (e.g., cGAN [11, 14, 40]). In addition, generative neural networks have

not been explored to assist human mobility response estimation at fine-scale in this COVID-19 pandemic.

Generative Adversarial Networks (GANs). GANs were proposed by Reference [14] and have achieved great performance in image generation domain, including image-to-image translation [17], image super-resolution [23], and text-to-image synthesis [29]. Despite the success, a critical issue for GANs is known to be the instability and sensitivity to the choices of hyper-parameters in learning process. Several works have attempted to address the GANs training problem and improved the stability by designing new network architectures [19], modifying the learning objectives and dynamics [25], and adding regularization methods to obtain stable gradients [6]. Besides image generation, recently, deep graph generative adversarial structure has been developed based on the concept of unsupervised learning. Existing architectures build upon the generative model, including GraphVAE [30] and GraphGAN [32], and they achieved good performance. However, the generative model has not been applied to estimating the human mobility problem as well as other human-related movement research.

8 CONCLUSIONS AND FUTURE WORK

We proposed a conditional COVID-GAN+ to estimate/generate real-world human mobility responses to assist policymaking during staged reopening in the face of the COVID-19 pandemic. In our prior work, we proposed a COVID-GAN that integrated various features, including contextual features, COVID-19 statistics, and policies from multi-sources. In addition, the generator incorporated a domain-constraint correction layer to reduce spurious results and the difficulty of learning efficiently. Experimental results showed that COVID-GAN can well mimic real-world human mobility responses with limitations in estimation outliers and does not consider spatial heterogeneity. Addressing these issues, in this article, we propose a COVID-GAN+ that tackles the above two limitations. We quantify the spatial heterogeneity and construct a spatial feature condition; in addition, we introduce an environmental constraint layer to force the model to learn the mobility changes to reduce the difficulty of estimating outliers. The experiment results show that the proposed model significantly improves the solution quality information.

In future work, we plan to explore spatial variability aware formulations of COVID-GAN as well as other measures of human mobility response and efficiently enforceable domain constraints with epidemiologists. Furthermore, the current work was limited to a single city. We will include more cities to improve the model capacity in transferring between different places. To achieve this goal, we will explore the meta-learning approaches to learn a shared knowledge across different cities and adapt to a new city with limited data. Also, we will consider distributed frameworks [27] for better scalability during data integration and model training.

ACKNOWLEDGMENTS

We thank SafeGraph Inc. (www.safegraph.com) for providing free access to Boston POI data, including Core Places, Geometry, and Weekly Places Patterns for this research.

REFERENCES

- [1] Centers for Disease Control and Prevention. 2020. Retrieved from https://www.cdc.gov/.
- [2] SafeGraph. 2020. Retrieved from https://www.safegraph.com/.
- [3] Luc Anselin. 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27, 2 (1995), 93-115.
- [4] Han Bao, Xun Zhou, Yingxue Zhang, Yanhua Li, and Yiqun Xie. 2020. COVID-GAN: Estimating human mobility responses to COVID-19 pandemic through spatio-temporal conditional generative adversarial networks. In Proceedings of the 28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL'20). Association for Computing Machinery, New York, NY, 273–282. https://doi.org/10.1145/3397536.3422261

27:22 H. Bao et al.

[5] Patrick Bryant and Arne Elofsson. 2020. Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. Retrieved from https://www.medrxiv.org/content/10.1101/2020.04.13.20063644v2.

- [6] Tong Che, Yanran Li, Athul Paul Jacob, Yoshua Bengio, and Wenjie Li. 2016. Mode regularized generative adversarial networks. Retrieved from https://arXiv:1612.02136\/.
- [7] Matteo Chinazzi, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun et al. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368, 6489 (2020), 395–400.
- [8] Hyunghoon Cho, Daphne Ippolito, and Yun William Yu. 2020. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. Retrieved from https://arXiv:2003.11511\forall.
- [9] Emre Eftelioglu, Shashi Shekhar, Dev Oliver, Xun Zhou et al. 2014. Ring-shaped hotspot detection: A summary of results. In *Proceedings of the IEEE International Conference on Data Mining*. IEEE, 815–820.
- [10] Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. SIGSPATIAL Special 12, 1 (2020), 16–26.
- [11] Jon Gauthier. 2014. Conditional generative adversarial nets for convolutional face generation. Technical report, Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, 5 (2014), 2.
- [12] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3656–3663.
- [13] Arthur Getis. 2008. A history of the concept of spatial autocorrelation: A geographer's perspective. *Geogr. Anal.* 40, 3 (2008), 297–309.
- [14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. MIT Press, 2672–2680.
- [15] Jayant Gupta, Yiqun Xie, and Shashi Shekhar. 2020. Towards spatial variability aware deep neural networks (SVANN): A summary of results. In Proceedings of the 1st ACM SIGKDD Workshop on Deep Learning for Spatiotemporal Data, Applications, and Systems.
- [16] Arthur E. Hoerl and Robert W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 1 (1970), 55–67.
- [17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1125–1134.
- [18] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. 2017. Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29, 10 (2017), 2318–2331.
- [19] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4401–4410.
- [20] Amin Vahedian Khezerlou, Xun Zhou, Lufan Li, Zubair Shafiq, Alex X. Liu, and Fan Zhang. 2017. A traffic flow approach to early detection of gathering events: Comprehensive results. ACM Trans. Intell. Syst. Technol. 8, 6 (2017), 1–24.
- [21] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Louis Du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368, 6490 (2020), 493–497.
- [22] Solomon Kullback. 1997. Information Theory and Statistics. Courier Corporation.
- [23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 4681–4690.
- [24] Yan Li, Pratik Kotwal, Pengyue Wang, Y. Xie, S. Shekhar, and W. Northrop. 2020. Physics-guided energy-efficient path selection using on-board diagnostics data. *ACM Trans. Data Sci.* 1, 3 (2020), 1–28.
- [25] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. 2017. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision. 2794–2802.
- [26] Zheyi Pan, Zhaoyuan Wang, Weifeng Wang, Yong Yu, Junbo Zhang, and Yu Zheng. 2019. Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2683–2691.
- [27] Sushil K. Prasad, Danial Aghajarian, Michael McDermott, et al. 2017. Parallel processing over spatial-temporal datasets from geo, bio, climate and social science communities: A research roadmap. In *Proceedings of the IEEE International Congress on Big Data (BigData'17)*. IEEE, 232–250.

- [28] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 779–788.
- [29] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative adversarial text to image synthesis. Retrieved from https://arXiv:1605.05396\/.
- [30] Martin Simonovsky and Nikos Komodakis. 2018. Graphvae: Towards generation of small graphs using variational autoencoders. In *Proceedings of the International Conference on Artificial Neural Networks*. Springer, 412–422.
- [31] Amin Vahedian, Xun Zhou, Ling Tong, W. Nick Street, and Yanhua Li. 2019. Predicting urban dispersal events: A two-stage framework through deep survival analysis on mobility data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5199–5206.
- [32] Hongwei Wang, Jia Wang, Jialin Wang, Miao Zhao, Weinan Zhang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2017. Graphgan: Graph representation learning with generative adversarial nets. Retrieved from https://arXiv:1711.08267V.
- [33] Zheng Wang, Kun Fu, and Jieping Ye. 2018. Learning to estimate the travel time. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18). Association for Computing Machinery, New York, NY, 858–866. https://doi.org/10.1145/3219819.3219900
- [34] Yiqun Xie, Rahul Bhojwani, Shashi Shekhar, and Joseph Knight. 2018. An unsupervised augmentation framework for deep learning based geospatial object detection: A summary of results. In Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 349–358.
- [35] Yiqun Xie, Jiannan Cai, Rahul Bhojwani, Shashi Shekhar, and Joseph Knight. 2020. A locally constrained yolo framework for detecting small and densely distributed building footprints. *Int. J. Geogr. Info. Sci.* 34, 4 (2020), 777–801.
- [36] Jiangye Yuan. 2017. Learning building extraction in aerial scenes with convolutional networks. *IEEE Trans. Pattern Anat. Mach. Intell.* 40, 11 (2017), 2793–2798.
- [37] Zhuoning Yuan, Xun Zhou, and Tianbao Yang. 2018. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 984–992.
- [38] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence.*
- [39] Xin Zhang, Yanhua Li, Xun Zhou, and Jun Luo. 2019. Unveiling taxi drivers' strategies via cGAIL: Conditional generative adversarial imitation learning. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'19)*. IEEE, 1480–1485.
- [40] Yingxue Zhang, Yanhua Li, Xun Zhou, Xiangnan Kong, and Jun Luo. 2019. TrafficGAN: Off-Deployment traffic estimation with traffic generative adversarial networks. In Proceedings of the IEEE International Conference on Data Mining (ICDM'19). IEEE, 1474–1479.

Received November 2020; revised June 2021; accepted August 2021