MULTI-TASK ANTI-CAUSAL LEARNING FOR RECONSTRUCTING URBAN EVENTS FROM RESIDENTS' REPORTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Many real-world machine learning tasks are anti-causal: they require inferring latent causes from observed effects. In practice, we often face multiple related tasks where the structural dependencies are a hybrid of task-invariant and task-specific mechanisms. We propose Multi-Task Anti-Causal learning (MTAC), a framework for estimating causes from outcomes and confounders by explicitly exploiting such cross-task invariances. MTAC learns a structural equation model (SEM) that factorizes the outcome-generation process into (i) a task-invariant mechanism and (ii) task-specific mechanisms via a shared backbone with task-specific deviations. Building on the learned forward model, MTAC performs maximum A posteriori (MAP) based inference to reconstruct causes by jointly optimizing latent mechanism variables and cause magnitudes under the learned structural model. We evaluate MTAC on the application of urban event reconstruction from resident reports, spanning three tasks: parking violations, abandoned properties, and unsanitary conditions. On real-world data collected from Manhattan and Newark, MTAC consistently improves reconstruction accuracy over strong baselines, achieving up to 33.04% MAE reduction and demonstrating the benefits of learning transferable mechanisms across tasks.

1 INTRODUCTION

Many real-world inference problems require estimating latent causes from observed effects, a setting commonly referred to as anti-causal learning. Examples include medical diagnosis (Ahsan et al., 2022), image classification (Machado et al., 2021), and urban event reconstruction (Zhang et al., 2022). In these problems, observations are generated through underlying mechanisms that transform causes into observable outcomes. When parts of these mechanisms remain invariant across multiple scenarios or tasks, exploiting such invariance can improve anti-causal inference.

Urban events, such as illegal parking and unsanitary issues, trigger residents' reports through digital platforms (e.g., 311 requests (NYC Open Data)). These reports are observations filtered through human decisions, which are shaped by many factors across various types of urban events, including neighborhood socioeconomic status (SES), access to technology, digital literacy etc. (Rotman et al., 2014; Mannarini et al., 2010; Kontokosta et al., 2017). Consequently, observed reports are not direct measurements of urban events but outcomes filtered through heterogeneous reporting behaviors. Recovering the underlying event distribution from such biased observations therefore requires modeling the report-generation process itself. We treat the reporting behavior itself as an unobserved mediator between event occurrence and the observed report (VanderWeele, 2015). This mediator is shaped by SES factors of the residents.

Identifying the true causal structure of residents' reporting behavior is inherently difficult, as it involves psychology and social science theory in the complex urban environment. Therefore, we do not aim to construct precise causal models for reporting behavior in this paper. However, the underlying causally-inspired relation structures serve as stronger proxies for reconstructing urban events. Specifically, the same observed report count may arise from fundamentally different combinations of true event incidence and reporting behavior. For example, a neighborhood with frequent events but low reporting activity can generate the same number of reports as a neighborhood with fewer

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

events but high reporting activity. Correlational models (Deng et al., 2022; Wang et al., 2025a; Butt et al., 2025a; Zhou et al., 2024b; Luan et al., 2024a) cannot distinguish between these explanations because they do not explicitly represent the report-generation process. By explicitly modeling how SES and event occurrence jointly generate reports through the reporting behavior, a structural dependency provides an inductive bias that helps disentangle reporting bias from true event incidence.

Moreover, reporting behavior exhibits partial invariance across event types. The reporting behavior on different events undergoes a similar psychological path, such as social identity (Mannarini et al., 2010) and self-efficacy (Rotman et al., 2014). For example, highly educated residents tend to have greater confidence in the effectiveness of formal complaints; consequently, they are more inclined to report problems irrespective of the specific event category (Kontokosta et al., 2017). Meanwhile, the reporting behavior also depends on the event itself. For example, residents may report events that they believe agencies will act on quickly (e.g., sanitation complaints) at higher rates than events with historically slow responses (e.g., abandoned property intake). Existing prediction methods (Agostini et al., 2024; Butt et al., 2025b; Zhou et al., 2024a; Luan et al., 2024b) treat each event type independently and therefore fail to exploit such shared mechanisms.

Exploiting these shared mechanisms requires addressing two challenges. First, the causal effects are generated by both task-invariant and task-specific mechanisms. In the example of urban event reporting, the reporting behavior of different types of events undergoes a similar psychological path, while also varies depending on the urgency and significance of the event itself. Therefore, it is critical to disentangle and learn the cross-task invariance. Second, even if the underlying mechanism is learned, the ultimate objective remains anti-causal: inferring latent event occurrence from observed reports. This requires jointly reasoning about unobserved causes and latent mechanism variables under the learned causal structure. From an anti-causal learning perspective, the partially invariant reporting mechanism provides additional structure that can make this problem more tractable.

To address these challenges, we propose a multi-task anti-causal learning framework, *MTAC*. To disentangle task-invariant and task-specific mechanisms, we impose a multi-task structural equation model (SEM) that explicitly represents the mechanism variables as latent factors and decomposes their generative process into task-invariant and task-dependent components, parameterized by separate neural network components. To address the second challenge, we develop a MAP-based inference algorithm that estimates the causes by jointly optimizing the causes and the shared mechanism variables under the imposed structural model.

We evaluate *MTAC* with the urban event reconstruction problem for three types of urban events: abandoned properties, illegally-parked vehicles, and sanitary issues collected through digital platforms (e.g., 311 requests (NYC Open Data)) from New York City and Newark.

Our contributions are summarized as follows:

- We propose MTAC, a structured multi-task SEM whose shared backbone captures the invariant SES to reporting-preference mechanism and whose task-specific heads capture event-specific reporting effects, paired with a MAP-based inversion algorithm that jointly recovers true event counts and latent reporting preferences.
- We formulate urban event reconstruction as a multi-task anti-causal inference problem in which observed reports arise from true events filtered through partially shared reporting mechanisms influenced by SES and latent reporting preferences.
- We evaluate MTAC on real-world data from Manhattan (parking violations, unsanitary conditions) and Newark (abandoned properties), demonstrating up to 34.61% MAE reduction over strong anti-causal and multi-task baselines.

2 Related Works

2.1 Multi-Task Learning and Multi-Task Causal Discovery

Multi-task learning (MTL) improves sample efficiency and generalization by sharing representations across related tasks while preserving task-specific components when necessary (Caruana, 1997; Tang et al., 2020). This shared-private design principle is particularly relevant when tasks exhibit both common drivers and task-specific dynamics. Separately, a growing line of work studies *multi-*

task or *multi-environment* causal discovery, aiming to recover causal graphs by pooling data from multiple related distributions or leveraging invariance across tasks/environments (Mooij et al., 2020; Huang et al., 2020; Yang et al., 2025). These methods mainly focus on identifying causal structures (graph recovery) under distribution shifts or heterogeneous datasets. In contrast, our primary goal is to estimate the cause given the outcome and confounders under a multi-task scenario.

2.2 Anti-causal Learning and causal prediction

Predicting causes from observed outcomes is fundamentally different from predicting outcomes from causes. In anticausal settings, accurate inference typically requires a generative model that captures how causes produce effects, enabling posterior inference over latent causes given observed effects (Schölkopf et al., 2012). Related problems appear in diagnosis, fault localization, and causal abduction, where one seeks likely explanations of observations under a structural causal model (Louizos et al., 2017; Pawlowski et al., 2020; Zhang et al., 2020). Nevertheless, much of this literature either (i) does not explicitly leverage the observed outcome information in a structured causal measurement setting, or (ii) relies on strong assumptions such as a observable mediator independent of confounders (Mao & Little, 2025). Our work targets an anti-causal estimation specific to urban event calibration: reports are outcomes generated from events through a latent reporting-preference mechanism. To address this, we perform maximum a posteriori (MAP) inference of event counts under a learned causal measurement model, effectively "inverting" the forward SEM to reconstruct the event distribution from biased reports.

2.3 Urban Event Prediction from Urban Service Data

A large body of work studies urban event prediction, including crime incidents (Butt et al., 2025b; Zhou et al., 2024a; Wang et al., 2025b), parking violations (Luan et al., 2024b), flooding (Agostini et al., 2024), and other localized urban disorders, from spatiotemporal signals and urban service data such as calls-for-service and 311-style complaints. Common modeling paradigms include spatiotemporal statistical models, point processes, and, more recently, deep learning approaches such as recurrent models and graph-based spatiotemporal networks that incorporate neighborhood context, mobility, and built-environment features (Butt et al., 2025b; Zhou et al., 2024a; Luan et al., 2024b). These methods have demonstrated strong predictive performance when the observed event logs are a reliable proxy of true underlying occurrences. However, for many civic issues, the observed "event" records are themselves partially generated by resident reporting behavior. Consequently, purely predictive models that treat reported records as ground truth may produce systematically distorted estimates of the event distribution, especially across regions with different socioeconomic status profiles.

3 Problem Definition

We aim to estimate causes from outcomes and confounders across K related tasks. In task k, let X_k and Y_k denote the cause and the outcome, and let $\mathbf{Z} = \{Z_l\}_{l=1}^{L}$ denote the observed confounders. X_k affects Y_k through M latent mechanism variables $\mathbf{W}^k = \{W_i^k\}_{i=1}^{M}$ as mediators, which are difficult to observe directly. Following Figure 1, the generative process of task k is

$$\begin{aligned} X_k &\sim p(X_k \mid \mathbf{Z}; \phi_k^X), \\ \mathbf{W}^k &\sim p(\mathbf{W}^k \mid \mathbf{Z}, X_k; \theta_k), \\ Y_k &\sim p(Y_k \mid \mathbf{W}^k; \phi_k^Y). \end{aligned} \tag{1}$$

For each task k, we aim to estimate X_k given Y_k and $\mathbf{Z}$, i.e., $p(X_k \mid Y_k, \mathbf{Z})$.

The process rests on two structural assumptions. **(A1) Complete mediation.** X_k and $\mathbf{Z}$ affect Y_k only through $\mathbf{W}^k$; there are no direct edges $X_k \to Y_k$ or $\mathbf{Z} \to Y_k$.

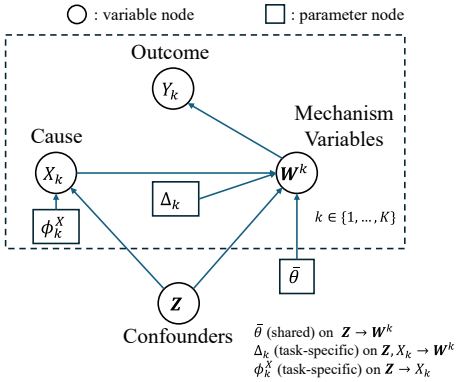


Figure 1: Multi-task causal graph with shared causal mechanism.

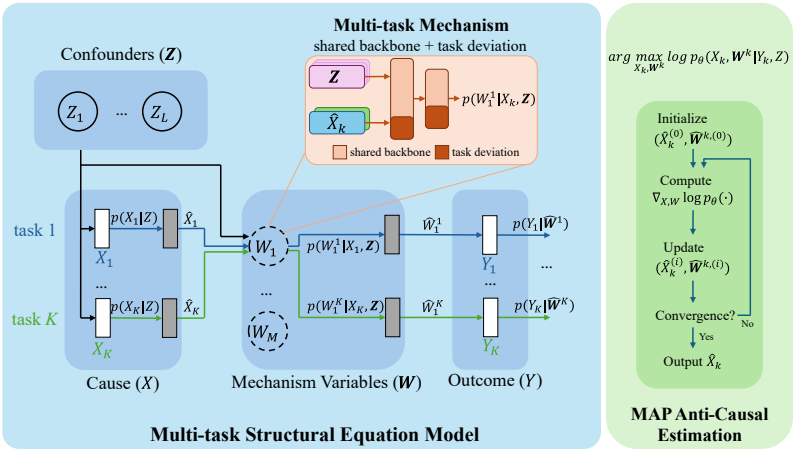


Figure 2: MTAC Framework. The white nodes represent parameterized deterministic neural networks, gray nodes represents drawing samples from the respective distribution. The left panel shows the multi-task structural equation model, in which the mechanism variables W and confounders Z are shared across tasks while causes X_k and outcomes Y_k are task-dependent. The orange box details the multi-task mechanism module for a single mediator W_1. The right panel illustrates the MAP inference procedure.

(A2) Shared mechanism parameters. The mechanism parameters decompose as $\theta_k = \bar{\theta} + \Delta_k$, where $\bar{\theta}$ is shared across tasks and Δ_k is sparse in the parameters acting on $\mathbf{Z}$. That is, $\bar{\theta}$ represents the task-invariant component in the mechanism, while Δ_k represents the task-dependent component.

Urban event reconstruction is a natural instance of this setting. A report is filed only when a resident decides to file it, so events and SES can affect report counts only through this reporting act, which motivates (A1). The way SES shapes the reporting act reflects properties of residents rather than of the event type, so we expect the parameters acting on $\mathbf{Z}$ to be largely shared across event types, while the response to the event itself is task-dependent, which motivates (A2). More details are in Appendix A.1.

4 Multi-task Structural Model

To estimate X_k from Y_k for task k, it is crucial to model the generation process of Y_k, which is a hybrid of task-invariant and task-dependent mechanisms. Given the heterogeneous datasets for each task, joint training across tasks provides an opportunity to learn a more generalizable and accurate task-invariant structural dependency. Therefore, we propose MTAC to model the generative mechanism of Y across multiple tasks. Figure 2 illustrates the framework of MTAC. As shown on the left side of Figure 2, a multi-task SEM is constructed to represent the structural dependencies based on Section 3. The cause and outcome in each task are modeled as separate variables. The mechanism variables and confounders are modeled as shared variables across tasks. The conditional distribution of each variable is modeled with a neural network module, represented as white nodes. Drawing samples from distributions is represented with gray nodes. As shown in the orange box in Figure 2, for the mechanism variable W_1, we explicitly capture the task-agnostic dependencies from $\mathbf{Z}$ with a task-shared backbone and the task-dependent dependencies from X_k with task deviation module. To address the anti-causal estimation based on the learned SEM, we employ a MAP-based inference algorithm to estimate the cause X_k, as shown on the right side of Figure 2.

4.1 Multi-task Structural Model

We model the structural dependencies with structural equation models. Specifically, the distribution of a variable $\mathbf{n}_i$ conditional on its parents pa_i is modeled as:

$$p(\mathbf{n}_i|\text{pa}_i) \sim \mathcal{N}(f_i^{\mu}(\text{pa}_i;\theta_i), f_i^{\sigma}(\text{pa}_i;\theta_i)). \tag{2}$$

$f_i^{\mu}(\cdot)$ and $f_i^{\sigma}(\cdot)$ fit the mean and variance of $p(\mathbf{n}_i|\text{pa}_i)$ respectively. $f_i^{\mu}(\cdot)$ and $f_i^{\sigma}(\cdot)$ are fitted using neural networks with parameter θ_i. For variables X and Y, $f_i^{\mu}(\cdot)$ and $f_i^{\sigma}(\cdot)$ are modeled with one

multi-layer perceptron (MLP) networks with two heads for μ and σ respectively. Particularly, to explicitly model the task-agnostic and task-dependent dependencies on the mechanism variables $\mathbf{W}$, a multi-task mechanism module is constructed. Details are discussed in Section 4.2.

The graph structure is constructed according to Figure 1, where acyclic constraint is naturally satisfied. Variable $\mathbf{n}_i$'s parents pa_i is identified with the first layer of $f_i(\cdot)$. Specifically, let $\boldsymbol{\theta}_i^{(1)}$ represent the parameter matrix of the first layer, variable $\mathbf{n}_j$ is $\mathbf{n}_i$'s parent only when the $\boldsymbol{\theta}_i^{(1)}$ have sufficiently significant path connecting $\mathbf{n}_j$ to the second layer, i.e., $||\boldsymbol{\theta}_{i;[j,:]}^{(1)}|| > c$ where c represents a threshold.

4.2 Multi-task Mechanism Module

The mechanism variables are generated under both task-dependent effects and task-shared effects. To model both structural dependencies, we construct a multi-task SEM for each mechanism variable, illustrated in Figure 3. For mechanism variable $\mathbf{W}_i$, all confounders $\mathbf{Z}$ and the cause X_k are supplied to a single network whose parameters at the first G layers are decomposed into a task-shared component and a task-dependent deviation (Evgeniou & Pontil, 2004). For layer $g \leq G$, we have:

$$\theta_{i,k}^{(g)} = \bar{\theta}_i^{(g)} + \Delta_{i,k}^{(g)}, \tag{3}$$

As such, that the mechanism variable is generated as

$$\mathbf{W}_i^k \sim p(\mathbf{W}_i^k \mid \mathbf{Z}, X_k;\ \bar{\theta}_i, \Delta_{i,k}). \tag{4}$$

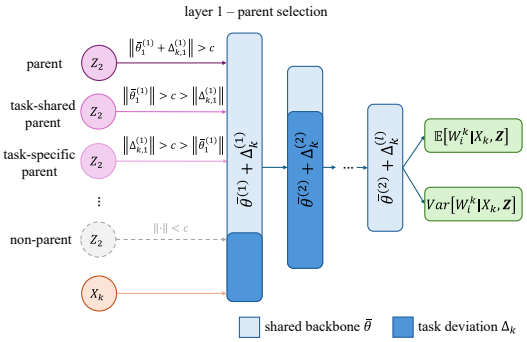


Figure 3: Multi-task SEM for mechanism variable $\mathbf{W}_i$. The norm of first-layer weights decides whether input variables are parents.

During training, the larger penalty is applied on the sparsity of $\Delta_{i,k}$ than $\bar{\theta}_i$. Details are discussed in Section 4.4. Since $\bar{\theta}_i$ is shared by all tasks while $\Delta_{i,k}$ only applies to task k, a variable that matters for every task is cheaper to encode once in $\bar{\theta}_i$ than separately in each $\Delta_{i,k}^{(1)}$, whereas a variable that matters for a single task is cheaper to encode in that task's deviation alone. This yields a structural disentanglement of the mechanism into an invariant and task-dependent component.

4.3 MAP Anti-Causal Estimation

The proposed forward structural model enables the estimation of the cause X_k of all tasks from its parental variables directly. However, such naive estimation does not leverage the information of the outcome Y_k. Therefore, we develop a MAP-based inference algorithm to estimate the cause by leveraging the outcome information. For task k, the value of X_k can be estimated with its posterior

$$(\hat{X}_k, \hat{\mathbf{W}}^k) = \max_{X_k, \mathbf{W}^k} p_\theta(X_k, \mathbf{W}^k | Y_k, \mathbf{Z}), \tag{5}$$

where θ represents the parameter of the SEM. Since $p_\theta(X_k, \mathbf{W}^k | Y_k, \mathbf{Z}) \propto p_\theta(X_k, Y_k, \mathbf{W}^k, \mathbf{Z})$, it is equivalent to solve

$$(\hat{X}_k, \hat{\mathbf{W}}^k) = \arg\max_{X_k, \mathbf{W}^k} \log p_\theta(X_k, Y_k, \mathbf{W}^k, \mathbf{Z}). \tag{6}$$

Notably, in equation (6), we also optimize over the mechanism variables $\mathbf{W}^k$ because they are latent and uncertain. Updating them makes the inference more stable. $p_\theta(X_k, Y_k, \mathbf{W}^k, \mathbf{Z})$ can be directly estimated with the forward structural model through the causal order as

$$\log p_\theta(\mathbf{n}) = \sum_i \log p_\theta(\mathbf{n}_i | \text{pa}_i). \tag{7}$$

Thus, equation (6) can be solved with gradient-based optimizers, such as Adam (Kingma & Ba, 2014), since the SEM is constructed using differentiable neural networks. We first initialize X_k as

Algorithm 1 MAP-based event inference

Input: Observation of outcome Y_k and confounders $\mathbf{Z}$
Output: Estimation of cause X_k
Initialize $X_k = \max p_\theta(X_k|\mathbf{Z})$.
Initialize $\mathbf{W} = \max p_\theta(\mathbf{W}|X_k, \mathbf{Z})$.
repeat
  Compute joint log probability density $\log p_\theta(\mathbf{n}) = \sum_i \log p_\theta(\mathbf{n}_i|\text{pa}_i)$ with SEM θ.
  Update X_k and $\mathbf{W}$ to minimize $-\log p_\theta(\mathbf{n})$ with gradient descent method.
until convergence

$p_\theta(X_k|\text{pa}(X_k))$ by inferring the structural model. Similarly, mechanism variables $\mathbf{W}^k$ are initialized by inferring the SEM as well. Then, X_k and $\mathbf{W}^k$ are optimized by solving equation (6) using gradient descent methods, while θ is frozen. The procedure is illustrated in Algorithm 1.

This event inference algorithm bridges the coupled multi-task anti-causal estimation and the shared mechanism. The mechanism variables $\mathbf{W}^k$ are generated jointly by the X_k and $\mathbf{Z}$. As a result, the same Y_k can be explained by either (1) a high dependency effect from X_k with a low effect from $\mathbf{Z}$, or (2) a lower effect from X_k with a higher effect of $\mathbf{Z}$. Taking an example of the urban event reconstruction problem, the same amount of reports can indicate a high amount of urban events with low reporting willingness, or a low amount of events with high reporting willingness. The proposed multi-task structural model supplies a transferable prior on latent mechanism variables, and the MAP inference algorithm leverages it to disentangle reporting bias from true event frequency. Equation (5) can be decomposed to

$$\begin{aligned}\arg\max_{X_k, \mathbf{W}^k} \log p_\theta(Y_k|\mathbf{W}_k) + \log p_\theta(\mathbf{W}^k|X_k, \mathbf{Z}) \\ + \log p_\theta(X_k|\mathbf{Z}).\end{aligned} \tag{8}$$

The first term represents the reporting model, which forces the inferred $\mathbf{W}^k$ to explain the observed outcome Y_k. The second term represents the shared multi-task prior, which regularizes $\mathbf{W}^k$ toward the value learned jointly across multiple tasks. The third term represents the task-dependent cause prior, which prevents degenerate event estimates that fit Y_k only by inflating X_k. As such, solving equation (5) balances the task-dependent and task-shared effects.

4.4 TRAINING

The learning parameters of MTAC are the network parameters θ of all SEMs. The graph structure is fixed by the causal order in Figure 1, so no adjacency matrix is optimized and no acyclicity constraint is required. Instead, the parents of each variable are recovered from the first-layer weights of its SEM, and the loss is designed so that this readout is meaningful. The loss consists of four terms:

$$\mathcal{L} = \mathcal{L}_{\text{nll}} + \mathcal{L}_{\text{sp}} + \lambda_{\text{div}}\mathcal{L}_{\text{div}} + \lambda_{\text{wd}}\mathcal{L}_{\text{wd}}. \tag{9}$$

Negative log likelihood. For variable $\mathbf{n}_i$, the negative log likelihood loss is

$$\mathcal{L}^i_{\text{nll}} = -\log p_\theta(\hat{\mathbf{n}}_i \mid \text{pa}_i), \tag{10}$$

which can be computed directly from the SEM in Eq. (2) without MAP inference. The confounders $\mathbf{Z}$ are observed exogenous variables and are conditioned on rather than modeled and $\mathbf{W}$ are latent, so no SEM is fitted for them and the overall loss is $\mathcal{L}_{\text{nll}} = \sum_{\mathbf{n}_i \notin \mathbf{Z} \bigcup \mathbf{W}} \mathcal{L}^i_{\text{nll}}$.

Sparsity of first-layer weights. As discussed in Section 4.1, variable $\mathbf{n}_j$ is identified as a parent of $\mathbf{n}_i$ when the corresponding row of the first-layer parameter matrix has a sufficiently large norm. To make this readout informative, we penalize the ℓ_2 norm of each input row, which drives all weights leaving an irrelevant input to zero simultaneously. For the cause variables, whose SEMs are task-dependent, this gives

$$\mathcal{L}^X_{\text{sp}} = \lambda_s \sum_{i \in I_X} \sum_j \left\|\theta^{(1)}_{i,j,:}\right\|_2. \tag{11}$$

For the mechanism variables, the first-layer weights are decomposed into the task-shared component $\bar{\theta}^{(1)}_i$ and the task deviation $\Delta^{(1)}_{i,k}$, and the two are penalized separately:

$$\mathcal{L}^W_{\text{sp}} = \lambda_s \sum_{i \in I_{\mathbf{W}}} \sum_{j \in \mathbf{Z}} \left\|\bar{\theta}^{(1)}_{i,j,:}\right\|_2 + \lambda_d \sum_{i \in I_{\mathbf{W}}} \sum_k \sum_{j \in \mathbf{Z}} \left\|\Delta^{(1)}_{i,k,j,:}\right\|_2, \tag{12}$$

with $\mathcal{L}_{\text{sp}} = \mathcal{L}_{\text{sp}}^{X} + \mathcal{L}_{\text{sp}}^{W}$. We set $\lambda_d > \lambda_s$ so that cross-task sharing is the default and a task deviation is introduced only when the data require it. The columns corresponding to the cause X_k are excluded from $\mathcal{L}_{\text{sp}}^{W}$, since the dependency of the mechanism on the cause is task-dependent by construction.

Non-redundancy of mechanism variables. The mechanism variables are latent and share the same parents, so nothing in the likelihood prevents several of them from encoding the same function. We discourage this with a redundancy penalty on their conditional means. For a mini-batch of task k, let $\tilde{\mu}_i$ denote the batch-standardized value of $f_i^{\mu}(\text{pa}_i; \theta)$ for $i \in I_{\mathbf{W}}$, and let ρ_{ij} be the empirical correlation between $\tilde{\mu}_i$ and $\tilde{\mu}_j$. Following (Bardes et al., 2022), we penalize the off-diagonal correlations,

$$\mathcal{L}_{\text{div}} = \sum_{k} \sum_{i<j,\ i,j \in I_{\mathbf{W}}} \rho_{ij}^2. \tag{13}$$

This term encourages the mechanism variables to span complementary directions rather than collapsing onto a single latent factor.

Scale control. Under positively homogeneous activations, a network can scale down its first-layer weights and scale up a later layer without changing the function it computes, which would reduce $\mathcal{L}_{\text{sp}}$ at no cost to the fit and render the parent readout vacuous. We therefore apply weight decay to all parameters after the first layer,

$$\mathcal{L}_{\text{wd}} = \sum_{i} \sum_{l>1} \left\| \theta_i^{(l)} \right\|_F^2, \tag{14}$$

which fixes the scale at which the first-layer norms are comparable across variables.

5 Evaluation

In this section, we present the empirical results on the urban event reconstruction problem. These experiments intend to answer the following questions: **Q1.** Does MTAC improve the performance of urban event reconstruction? (§5.2) **Q2.** Is the task-invariant mechanism valid across tasks? (§5.4) **Q3.** How much does jointly learning the task-invariant mechanism improve cause estimation? (§5.3) Appendices A.4 $\sim$ A.6 answers several additional question to provide a more comprehensive evaluation on MTAC.

5.1 Evaluation Setting

We evaluate our design for the application of urban event reconstruction. Three types of urban events are considered: 1) parking violations, 2) abandoned properties, and 3) unsanitary conditions. The data is collected in New York City and City of Newark. Details are discussed in Appendix A.2.

To our knowledge, no prior method targets the joint setting of multi-task learning, anti-causal estimation, and continuous causes; consequently, we compare against the strongest available baselines from each adjacent line of work and adapt them to our setting where necessary.

BSM-UR (Agostini et al., 2024): A Bayesian spatial latent-variable model that recovers the underlying true event incidence from biased reports. BSM-UR does not model the causal mechanism of generating resident reports. It only considers a single type of event.

PLE (Tang et al., 2020): A multi-task learning architecture that stacks customized gate-control layers to progressively extract shared and task-dependent features via mixtures of shared/task experts. PLE does not model the causal relationship between inputs and outputs.

CEVAE (Louizos et al., 2017): A deep latent-variable causal model that uses a variational autoencoder to account for hidden confounding and enables counterfactual/causal-effect inference (and posterior inference over latent causes) from observational data. To adapt CEVAE to causes of continuous values, we replace the networks $q(y|t=1,x)$ and $q(y|t=0,x)$ with a network $q(y|t,x)$ that takes the cause t as input. To estimate the cause, the $p(t|z)$ network in the decoder is inferred.

TEDVAE (Zhang et al., 2021): A variational latent-factor causal model that infers hidden factors from observed covariates and disentangles them into instrumental, confounding, and risk components. To evaluate TEDVAE, we apply the PC algorithm (Spirtes et al., 2000) for causal discovery

Model	Parking Violation		Abandoned Property		Unsanitary Condition	
	MAE	MSE	MAE	MSE	MAE	MSE
MTAC	**0.2971**	**0.1616**	**0.4163**	**0.1910**	**0.3755**	**0.2501**
CEVAE	0.3235	0.1925	0.4867	0.2607	0.4091	0.2884
TEDVAE	0.4315	0.2636	0.4932	0.2583	0.4188	0.2700
BSM-UR	0.4437	0.2976	0.5438	0.3175	0.3982	0.3026
PLE	0.3337	0.2174	0.4533	0.2351	0.4101	0.2697

Table 1: MAE and MSE estimation for MTAC and baselines across three tasks.

Model	Parking Violation		Abandoned Property		Unsanitary Condition	
	MAE	MSE	MAE	MSE	MAE	MSE
Multi-task	**0.2971**	**0.1616**	**0.4163**	**0.1910**	**0.3755**	**0.2501**
Single-task	0.3693	0.2947	0.4974	0.2517	0.4486	0.2989

Table 2: Performance comparison for MTAC between multi-task training and single-task training.

and categorize the SES variables into these three groups. To adapt to the causes of continuous values, TEDVAE is modified in the same way as CEVAE. We infer the auxiliary classifier $q_{w_t}(t|\mathbf{z}_t, \mathbf{z}_c)$ for cause estimation. TEDVAE does not employ multi-task learning.

To ensure fairness from a data perspective, we train these baselines on the concatenation of all three task datasets, with task indicator as an extra feature. The event distribution prediction error is evaluated as the difference between the inferred value of the urban event count and the ground-truth. We measure the prediction error using MAE and MSE. To validate the contribution of the task-agnostic structural dependencies, we also compare the prediction error when trained separately on each task and when jointly trained across the three tasks. Moreover, we learn MTAC on two tasks and transfer the task-shared backbone to the other task to validate the task-invariant component of structural dependencies from the SES factors. Other results are presented in the Appendices A.4 $\sim$ A.6. Ablation studies are conducted to illustrate the contribution of each component in MTAC (Appendix A.4). Policy sensitivity analysis is conducted to understand MTAC's causal utility (Appendix A.5). A parameter sweep is conducted to understand the multi-path psychological model for reporting behavior (Appendix A.6).

5.2 Prediction Error

Table 1 presents MTAC's prediction error against other baselines. MTAC outperforms all baselines. MTAC reduces MAE relative to CEVAE, TEDVAE, and BSM-UR by 8.16%, 31.15%, and 33.04% on parking violations; 14.45%, 15.59%, and 23.45% on abandoned properties; and 8.21%, 10.34% and 5.70% on unsanitary conditions. The comparison on MSE is also consistent. This result demonstrates that MTAC learns a more accurate model because the task-agnostic mechanism is learned using data from multiple tasks. This jointly learned mechanism benefits the reconstruction of all three tasks. On the other hand, PLE does not model the causal relationship in report generation and only learns the association. On each dataset, MTAC reduces MAE by 10.97%, 8.16%, and 8.44% compared to PLE, respectively. MTAC also outperforms PLE in MSE. This highlights the importance of modeling the structural relationship to avoid spurious associations.

5.3 Multi-task v.s. Single-task

To further validate whether jointly learning the shared mechanism contributes to cause estimation, we train and test MTAC on each task separately. The comparison to jointly-trained MTAC is presented in Table 2. When jointly trained on all datasets, the prediction error of MTAC decreases compared to the model trained on each task separately. MAE and MSE decreased across all tasks: parking violation (0.0722, 0.1331), abandoned property (0.0811, 0.0617), and unsanitary condition (0.0731, 0.0488). This reduced prediction error in all three tasks demonstrates that jointly learning the task-agnostic structural dependencies leads to a more accurate model. When learned with single-task datasets, MTAC performed comparably to the baselines on the parking violation and abandoned property tasks and worse than the baselines on the unsanitary condition task. This is because the baselines are trained using all datasets. Even with models not designed for multi-task learning, the enlarged dataset already provides a performance gain.

Model	UC+PV → AP		UC+AP → PV		PV+AP → UC	
	MAE	MSE	MAE	MSE	MAE	MSE
Full fine-tuned	**0.4507**	0.2313	0.3475	**0.2502**	**0.3977**	**0.2656**
Deviation-only fine-tuned	0.4631	**0.2297**	**0.3407**	0.2586	0.4106	0.2694
Zero-shot	0.7702	0.4323	0.7699	0.5472	0.8965	0.7518
Single-task trained	0.4974	0.2517	0.3693	0.2947	0.4486	0.2989

Table 3: Prediction Error of transferred MTAC. The transferred models are trained on two tasks and transferred to the remaining task. PV: parking violation; AP: abandoned property; UC: unsanitary condition.

5.4 Task-invariant Mechanism Validation Across Tasks

To demonstrate that the structural dependencies of the SES on reporting preferences are invariant across urban event types, we also evaluated the reconstruction performance by transferring the model parameters to an unseen task. We conduct a three-fold evaluation. In each fold, we select two tasks and train an MTAC as a teacher model, while a student model is trained on the remaining task. The shared backbone $\bar{\theta}$ is transferred from the teacher model to the student model using several transfer learning paradigms:

Zero-shot: The student model is first trained on the target task. For all mechanism variables, we replace the student model's shared backbone $\bar{\theta}$ in the SEMs with that of the teacher model. Then, the student model is evaluated directly without further fine-tuning.

Deviation-only fine-tuning: Starting from the same teacher initialization, we transfer the parameters of $\bar{\theta}$ to the student model and freeze this shared module. We then train only the parameter of target-task deviation module Δ for mechanism variables and the complete SEMs for other variables) using the target-task training dataset while freezing $\bar{\theta}$.

Full fine-tuning: For the fully fine-tuned student model, $\bar{\theta}$ is also optimized after being transferred from the teacher model.

Table 3 presents the reconstruction error of each student model on each dataset. The performance of a single-task trained version of MTAC is also included as a baseline. In all three evaluations, both the deviation-only and fully fine-tuned student models outperform the single-task trained model. Particularly, the deviation-only fine-tuned model uses the exact backbone $\bar{\theta}$ learned from the other two tasks. It shows higher accuracy than the single-task trained model in all three folds. This validates that the structural dependencies from the confounders to the mechanism variables holds invariance across tasks. When transferring to the abandoned property task, the reduction in both MAE and MSE is most significant compared to the other tasks. Again, this is due to the limited size of the abandoned property dataset. Moreover, the zero-shot model performs poorly on all tasks due to heterogeneity across tasks. The full fine-tuned model and deviation-only fine-tuned model perform similarly in each fold. This is because the task-shared backbone captures the task-agnostic patterns and is updated only slightly during full fine-tuning. The average normalized L_2 distance between the task-shared backbone in fully fine-tuned and deviation-only fine-tuned model is 0.064, which represents a small parameter distance.

6 Conclusion

This work formulates the multi-task anti-causal learning problem and proposes MTAC to leverage the task-invariant structural depenencies for enhancing cause estimation. MTAC consists of a multi-task structural equation model while explicitly modeling the task-invariant and task-dependent structural dependencies of latent mechanism variables. The task-shared and task-dependent mechanisms are explicitly captured using separate neural network modules. A MAP-based inference algorithm is applied to invert the learned forward SEM for estimating the magnitude of causes from outcomes. Three urban event reconstruction tasks are evaluated as applications: parking violations, abandoned properties, and unsanitary conditions. Evaluations show that jointly learning the shared mechanism improves reconstruction accuracy compared with single-task training. The results also validate the shared structural dependencies across tasks.

7 AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on research methodology or experiments. We have not used generative AI tools to assist with translation, clean and reformat datasets, support qualitative and thematic data analysis, and interpret results. Generating synthetic data sets, formulating mathematical claims, providing critical ingredients for proving mathematical claims, and assisting in the writing of proofs are not applicable to this work. Additionally, we used generative AI tools for brainstorming, creating and editing software code, polishing paper writing to improve readability, and formatting references. We have reviewed all AI-assisted work. We have verified the code integrity and language usage of the paper. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

8 REPRODUCIBILITY STATEMENT

We have made several efforts to ensure the reproducibility of our results. The problem setting and the structural assumptions underlying MTAC are stated in Section 3. The model architecture, the MAP inference procedure (Algorithm 1), and the complete training objective with all loss terms are described in Sections 4.1–4.4. Appendix A.2 lists the data sources for all three tasks, the spatial and temporal aggregation, and the socioeconomic factors used (Table 4). Appendix A.3 reports the network architectures, hyperparameters, data splits, normalization, and hardware. All report and ground-truth data for the Manhattan tasks, as well as the census data, come from publicly available sources cited in Appendix A.2. The Newark abandoned-property records were obtained from the City of Newark; we describe their processing in Appendix A.2.

REFERENCES

Gabriel Agostini, Emma Pierson, and Nikhil Garg. A bayesian spatial model to correct underreporting in urban crowdsourcing. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i20.30190.

Md Manjurul Ahsan, Shahana Akter Luna, and Zahed Siddique. Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare*, 10(3):541, 2022. doi: 10.3390/healthcare10030541.

Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2022.

Umair Muneer Butt, Sukumar Letchmunan, Mubashir Ali, and Hafiz Husnain Raza Sherazi. Start: A spatiotemporal autoregressive transformer for enhancing crime prediction accuracy. *IEEE Transactions on Computational Social Systems*, 12(6):4650–4664, 2025a. doi: 10.1109/TCSS.2025.3550196.

Umair Muneer Butt, Sukumar Letchmunan, Mubashir Ali, and Hafiz Husnain Raza Sherazi. Start: A spatiotemporal autoregressive transformer for enhancing crime prediction accuracy. *IEEE Transactions on Computational Social Systems*, 12(6):4650–4664, 2025b. doi: 10.1109/TCSS.2025.3550196.

Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, July 1997. ISSN 1573-0565. doi: 10.1023/A:1007379606734. URL https://doi.org/10.1023/A:1007379606734.

City of New York. Parking violations issued - fiscal year 2024. NYC Open Data, 2024a. URL https://data.cityofnewyork.us/City-Government/Parking-Violations-Issued-Fiscal-Year-2024/pvqr-7yc4.

City of New York. Oath hearings division case status. NYC Open Data, 2024b. URL https://data.cityofnewyork.us/City-Government/OATH-Hearings-Division-Case-Status/jz4z-kudi/about_data.

Scott Cook, Samantha Zuhlke, and Robin Saywitz. Potholes, 311 reports, and a theory of heterogeneous resident demand for city services. *Policy Studies Journal*, 52:647–669, 06 2024. doi: 10.1111/psj.12540.

Songgaojun Deng, Huzefa Rangwala, and Yue Ning. Causality enhanced societal event forecasting with heterogeneous graph learning. In *2022 IEEE International Conference on Data Mining (ICDM)*, pp. 91–100, 2022. doi: 10.1109/ICDM54844.2022.00019.

Theodoros Evgeniou and Massimiliano Pontil. Regularized multi–task learning. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pp. 109–117, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138881. doi: 10.1145/1014052.1014067. URL https://doi.org/10.1145/1014052.1014067.

Biwei Huang, Kun Zhang, Mingming Gong, and Clark Glymour. Causal discovery from multiple data sets with non-identical variable sets. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10153–10161, Apr. 2020. doi: 10.1609/aaai.v34i06.6575. URL https://ojs.aaai.org/index.php/AAAI/article/view/6575.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Constantine Kontokosta, Boyeong Hong, and Kristi Korsberg. Equity in 311 reporting: Understanding socio-spatial differentials in the propensity to complain. *arXiv preprint arXiv:1710.02452*, 2017.

Constantine E. Kontokosta and Boyeong Hong. Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. *Sustainable Cities and Society*, 64:102503, 2021. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2020.102503. URL https://www.sciencedirect.com/science/article/pii/S2210670720307216.

Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6449–6459, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Dongming Luan, En Wang, Nan Jiang, Bo Yang, Yongjian Yang, and Jie Wu. A data-driven crowdsensing framework for parking violation detection. *IEEE Transactions on Mobile Computing*, 23 (6):6921–6935, 2024a. doi: 10.1109/TMC.2023.3331429.

Dongming Luan, En Wang, Nan Jiang, Bo Yang, Yongjian Yang, and Jie Wu. A data-driven crowdsensing framework for parking violation detection. *IEEE Transactions on Mobile Computing*, 23 (6):6921–6935, 2024b. doi: 10.1109/TMC.2023.3331429.

Gabriel Resende Machado, Eugênio Silva, and Ronaldo Ribeiro Goldschmidt. Adversarial machine learning in image classification: A survey towards the defender's perspective. *ACM Computing Surveys*, 55(1):Article 8, 2021. doi: 10.1145/3485133.

Terri Mannarini, Angela Fedi, and Stefania Trippetti. Public involvement: How to encourage citizen participation. *Journal of Community & Applied Social Psychology*, 20(4):262–274, 2010. doi: https://doi.org/10.1002/casp.1030. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/casp.1030.

Jianqiao Mao and Max A. Little. Mechanism learning: reverse causal inference in the presence of multiple unknown confounding through causally weighted gaussian mixture models, 2025. URL https://arxiv.org/abs/2410.20057.

Joris M. Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *J. Mach. Learn. Res.*, 21(1), January 2020. ISSN 1532-4435.

NYC Open Data. 311 Service Requests from 2020 to Present. NYC Open Data (Socrata dataset ID: erm2-nwe9). URL https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2020-to-Present/erm2-nwe9. Accessed: 2026-02-06.

Nick Pawlowski, Daniel C. Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

N. Andrew Peterson and Robert J. Reid. Paths to psychological empowerment in an urban community: Sense of community and citizen participation in substance abuse prevention activities. *Journal of Community Psychology*, 31(1):25–38, 2003. doi: https://doi.org/10.1002/jcop.10034. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcop.10034.

Dana Rotman, Jennifer Hammock, Jenny Preece, Derek Hansen, Carol Boston, Anne Bowser, and Yurong He. Motivations affecting initial and long-term participation in citizen science projects in three countries. 03 2014. doi: 10.9776/14054.

Bernhard Schölkopf, Dominik Janzing, Jonas Peters, Eleni Sgouritsa, Kun Zhang, and Joris Mooij. On causal and anticausal learning. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pp. 459–466, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.

SeeClickFix. Seeclickfix - newark, nj. https://seeclickfix.com/newark, 2024.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000.

Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems*, RecSys '20, pp. 269–278, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375832. doi: 10.1145/3383313.3412236.

U.S. Census Bureau. Explore census data (data.census.gov). Website. URL https://data.census.gov/. Accessed: 2026-02-01.

Tyler VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.

Zepu Wang, Xiaobo Ma, Huajie Yang, Weimin Lyu, Yang Liu, Peng Sun, and Sharath Chandra Guntuku. Uncertainty-aware crime prediction with spatial temporal multivariate graph neural networks. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2025a. doi: 10.1109/ICASSP49660.2025.10889685.

Zepu Wang, Xiaobo Ma, Huajie Yang, Weimin Lyu, Yang Liu, Peng Sun, and Sharath Chandra Guntuku. Uncertainty-aware crime prediction with spatial temporal multivariate graph neural networks. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2025b. doi: 10.1109/ICASSP49660.2025.10889685.

Dezhi Yang, Guoxian Yu, Jun Wang, Jinglin Zhang, and Carlotta Domeniconi. Causal discovery from shifted multiple environments. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, pp. 1751–1762, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712456. doi: 10.1145/3690624.3709247.

Mingyang Zhang, Tong Li, Yue Yu, Yong Li, Pan Hui, and Yu Zheng. Urban anomaly analytics: Description, detection, and prediction. *IEEE Transactions on Big Data*, 8(3):809–826, 2022. doi: 10.1109/TBDATA.2020.2991008.

Weijia Zhang, Lin Liu, and Jiuyong Li. Treatment effect estimation with disentangled latent factors. In *AAAI Conference on Artificial Intelligence*, 2020. URL https://api.semanticscholar.org/CorpusID:210943075.

Weijia Zhang, Lin Liu, and Jiuyong Li. Treatment effect estimation with disentangled latent factors. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):10923–10930, May 2021. doi: 10.1609/aaai.v35i12.17304. URL https://ojs.aaai.org/index.php/AAAI/article/view/17304.

Binbin Zhou, Hang Zhou, Weikun Wang, Liming Chen, Jianhua Ma, and Zengwei Zheng. Hdm-gnn: A heterogeneous dynamic multi-view graph neural network for crime prediction. *ACM Trans. Sen. Netw.*, May 2024a. ISSN 1550-4859. doi: 10.1145/3665141. Just Accepted.

Binbin Zhou, Hang Zhou, Weikun Wang, Liming Chen, Jianhua Ma, and Zengwei Zheng. Hdm-gnn: A heterogeneous dynamic multi-view graph neural network for crime prediction. *ACM Trans. Sen. Netw.*, May 2024b. ISSN 1550-4859. doi: 10.1145/3665141. Just Accepted.

A Appendix

A.1 A Case Study: Urban Event Reconstruction

Urban event reconstruction is an application where multi-task anti-causal learning being effective. Residents submit reports upon observing various urban events, such as parking violations, abandoned properties, and unsanitary sites. This procedure can be understood as a causal generation, where urban events cause residents' reports through a causal mechanism representing their reporting preferences, which are influenced by the residents' socioeconomic status (Kontokosta et al., 2017). Thus, the causal model for urban report generation is modeled as follows: 1) cause (X): number of urban events; 2) outcome (Y): number of resident reports; 3) confounders (Z): residents' SES; 4) mechanism variables (W): residents' reporting preference.

The reporting act is widely studied in social research (Kontokosta et al., 2017; Peterson & Reid, 2003). The SES factors that shape whether and how a resident files a report can be grouped into five broad categories: 1) financial, 2) educational attainment, 3) race and culture, 4) access to technology, and 5) social environment. For example, people with higher educational attainment often trust more that their complaints will be addressed (Cook et al., 2024). Social studies (Mannarini et al., 2010; Rotman et al., 2014) identify a corresponding set of psychological pathways through which SES influences the reporting behavior, such as social identity, trust in government, self-efficacy, emotion, and digital literacy. Motivated by this social-science evidence for a multi-dimensional reporting-behavioral representation, we construct multiple latent mechanism variables. Because ground-truth measurements of these latent factors are not available at scale in urban data, MTAC does not enforce or claim a one-to-one correspondence between individual W_i dimensions and named psychological variables. Individual dimensions are treated as unlabeled latent factors on the reporting-act pathway, and their contribution to the model is evaluated through end-to-end reconstruction of X rather than through direct semantic validation.

Notably, these psychological pathways are independent of the urban event type. Therefore, learning a unified reporting preference model across multiple tasks is more generalizable and may benefit each task. The causal effect of urban event occurrence is task-dependent, reflecting the heterogeneous reporting propensities of residents toward different categories of urban incidents (Kontokosta & Hong, 2021). As such, the multi-task causal mechanism design in MTAC would explicitly capture both the task-shared and task-dependent causal effects of residents' reporting behavior.

A.2 Datasets

We evaluate our design for the application of urban event reconstruction. Three types of urban events are considered: 1) parking violations, 2) abandoned properties, and 3) unsanitary conditions. For abandoned properties, we collect data from the City of Newark. The reports are collected from the SeeClickFix platform (SeeClickFix, 2024). This is an online platform where residents can submit complaints about various urban issues. We collect the labels of abandoned properties from the historical records of the City of Newark, which track and monitor abandoned properties. The data are collected from a system for the city of Newark to track and monitor the status of abandoned properties. When an abandoned property is identified by the city agents, a record is created in the system. Each record tracks the progress of handling an abandoned property, including inspections, rehabbing, and demolishing. It also includes information about when the properties are detected by the city. With detailed tracking of abandoned properties, we label the properties in the time domain to avoid mismatches between labels and multi-view data sources. Specifically, we assigned a detection time for each property, where no events after the time are collected for all views. We use

Category	Factors
Finance	mean income, unemployment rate, mortgage ratio, poverty rate, housing cost
Education Attainment	% less than high school, % high school, % bachelor or higher
Race & Culture	% Hispanic, % white, % black, % Asian
Access to Technology	% has computer, % has smartphone, % has internet
Social Environment	population, % multifamily, % owner occupied, % renter occupied, median year built, % room occupation ≥ 0.5, mobility rate

Table 4: An overview of the socioeconomic status factors.

Model	Parking Violation		Abandoned Property		Unsanitary Condition	
	MAE	MSE	MAE	MSE	MAE	MSE
Complete MTAC	**0.2971**	**0.1616**	**0.4163**	**0.1910**	**0.3755**	**0.2501**
w/o MAP	0.5841	0.6977	0.7841	0.6006	0.5957	0.7426
w/o Causality	0.3916	0.1811	0.4967	0.2842	0.4287	0.2938

Table 5: Ablation study. For MTAC without MAP, the cause is estimated directly from the forward causal model.

this information to count the number of abandoned properties during each time period. We collect the data from Jan. 2019 to May 2021 in Newark.

We collect data on parking violations and unsanitary issues from Manhattan in 2023. The residents' reports are collected from the NYC311 platform. The ground-truth of parking violations is collected from the database of parking violation tickets provided by the department of finance (City of New York, 2024a). The ground-truth of unsanitary conditions is collected from the hearing records of the city's administrative law court (City of New York, 2024b).

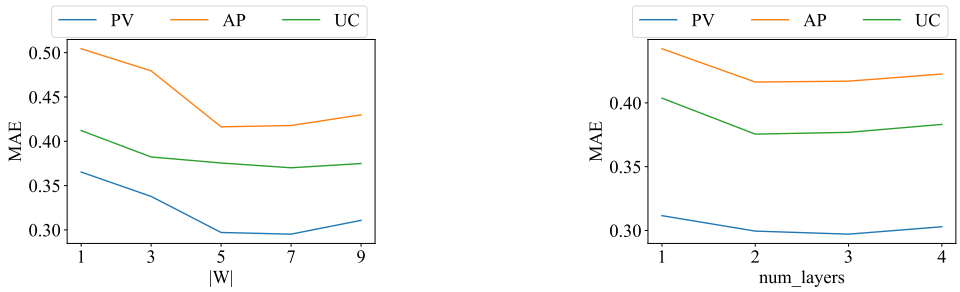
The cities are partitioned based on census tracts, and we collect the residents' socioeconomic status in each region from the census bureau database (U.S. Census Bureau). As a result, the city of Newark is partitioned into 88 regions, and Manhattan is partitioned into 322 regions. We count the frequency of urban events and reports every month. The socioeconomic status factors include 5 categories, as shown in Table 4. There are 22 factors collected, and the details of each factor can be found in (U.S. Census Bureau).

A.3 Implementation details

The multi-task SEM is implemented according to Figure 2. Each neural network block (the white nodes) is implemented using a 4-layer MLP with a hidden dimension of 64. For the mechanism variables, their generation is fitted with the module in Figure 3, where each neural network block is also implemented using an MLP with a hidden dimension of 64. As discussed in Section A.1, we set the number of mechanism variables as $|W| = 5$ to represent the psychological pathways that affect reporting behavior. For mechanism variables' SEMs, the first 2 layers are partitioned into the task-shared backbone and task-dependent deviation. Training parameters are set as: $\lambda_s = 10^{-3}$, $\lambda_d = 5 \cdot 10^{-3}$, $\lambda_{\text{div}} = 10^{-4}$, $\lambda_{\text{wd}} = 10^{-4}$ and $c = 0.5$. 80% of the data is used for training and validation, while the other 20% is used for evaluation. Z-score normalization is applied. Since BSM-UR, CEVAE, and TEDVAE do not employ multi-task learning, they are trained and evaluated on each task separately. A 10-fold experiment is conducted on a personal computer with an Intel 13700K CPU and an NVIDIA 4070 Ti SUPER GPU. The training takes 253.2 seconds on average for each trial.

A.4 Ablation Study

To understand the contribution of each component of MTAC, we conduct an ablation study. The results are presented in Table 5. We evaluate MTAC when the cause is estimated directly from the forward causal model instead of using the MAP-based inference algorithm. The prediction error increases significantly on all three tasks. On the abandoned property dataset, the MAE increases



(a) MAE against number of mechanism variables.

(b) MAE against deviation layers.

Figure 4: Sensitivity of reconstruction performance on model parameters.

by 254.49%. This is because the causal model is a forward model, which computes the distribution of variables given its parental variables. Direct cause estimation only uses the SES information of residents but not the reports, i.e., $p(X|Z)$ rather than $p(X|Z, Y)$. Therefore, the direct estimation is highly inaccurate without the strongest indicator Y.

To estimate the gain of modeling the causal structure compared to correlational relations, we evaluate MTAC without the causal structure. Specifically, a task-shared MLP module transfers SES variables to an embedding vector. For each task, a task-dependent head takes the embedding vector concatenated with the report count as input and generates the estimated event count. This model learns the correlative relation across multiple tasks, rather than modeling the causal structure of reporting behavior. As shown in Table 5, the complete MTAC achieves higher reconstruction accuracy in all tasks. By modeling the causal structure, MTAC separates the reporting bias and true event incidence through joint inference under the assumed structure. The gain in reconstruction accuracy highlights the significance of modeling causal structure.

A.5 Policy Sensitivity Analysis

To understand MTAC's causal utility, we also evaluate MTAC's performance under distribution shift. Specifically, we evaluate the prediction accuracy in regions with various SES profiles. We apply principal component analysis (PCA) to SES and stratify the test tracts into 4 SES quartiles based on the PC1 component. MTAC's reconstructed true event count achieves 2.2–29 times lower MAE than leveraging raw reports as event observations across every task and quartile. Across all three tasks and all four SES quartiles, MTAC consistently attains lower MAE than raw reports: for abandoned property, MTAC achieves 0.11–0.13 events/tract versus 2.2–3.7 for raw reports (ratio: 18.8–28.9×), with the largest gap in the lowest-SES quartile (Q1: 0.129 vs. 3.722, ratio 28.9×); for parking violation, MTAC achieves 13.3–14.6 versus 39.7–51.0 (ratio: 2.7–3.8×), with broadly stable MTAC MAE across quartiles (range: 1.3 events/tract) despite large variation in raw report counts (range: 11.3 events/tract); and for sanitary condition, MTAC achieves 2.4–3.4 versus 5.3–8.1 (ratio: 2.2–2.7×), where the absolute gap widens monotonically from high- to low-SES tracts (Q4: 2.367 vs. 5.277; Q1: 3.358 vs. 8.113). Raw reports show strongly SES-varying bias (under-counting by 6.8 events in low-SES vs. 2.6 in high-SES quartiles), while MTAC's worst-case bias across all tasks and quartiles is less than 3.4 events. This demonstrates causal utility beyond curve-fitting: MTAC recovers an estimator robust to confounder shift.

A.6 Hyperparameter Sensitivity

As inspired by (Mannarini et al., 2010; Rotman et al., 2014), there are multiple psychological paths that affect residents' reporting behavior, as discussed in Appendix A.1. To understand the dimonsionality the multi-path psychological mechanism, we evaluate MTAC while sweeping the number of mechanism variables. Figure 4a illustrates the reconstruction MAE performance accordingly. The reconstruction error is improved as the number of mechanism variables reaches 5, and stay stable afterwards. An overfitting phenomenon is observed with 9. The degradation in reconstruction loss with fewer mechanism variables indicates that the reporting preference is high-dimensional. It is a compound effect of multiple psychological paths. As such, it is necessary for MTAC to have

multiple mechanism variables to model reporting preferences. This result shows that there exist multi-dimensional psychological path of residents' reporting behavior in (Kontokosta et al., 2017; Mannarini et al., 2010; Peterson & Reid, 2003; Rotman et al., 2014).

Figure 4b demonstrates the reconstruction MAE with the number of layers being partitioned into task-shared backbone and task-dependent deviation. We implement the mechanism SEM module with 4 layers. Although only first layer serves parent selection, extending the deviation module to more layers shows reconstruction improvement across all tasks. This is because, besides the structural relation, the dependency effect is also task-dependent, and deviation module with only one layer is not representative enough to capture it.