International Journal of Safety and Security Engineering
Vol. 16, No. 10, October, 2026, pp. 2259-2273
Journal homepage: http://iieta.org/journals/ijsse

Architecture-Aware Operational Crowd-Safety Reasoning under Partial Observation

Abdullah Almohammadi

Department of Artificial Intelligence and Data Science, Taibah University, Madinah 41411, Saudi Arabia

Corresponding Author Email: aalmohamadi@taibahu.edu.sa

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ijsse.161006

Received: 16 August 2026
Revised: 23 September 2026
Accepted: 9 October 2026
Available online: 31 October 2026

Keywords:
architecture-aware reasoning, computer vision, crowd safety, egress sufficiency, partial observation, spatio-temporal graph neural network

ABSTRACT

Operational crowd safety depends on the interaction between pedestrian demand and architectural capacity, not on visible crowd density alone. Because substantial portions of real venues lie outside camera coverage, assessment becomes a reasoning problem under partial observation. This paper reformulates the problem as architecture-aware reasoning over a capacitated venue graph. A constructive non-identifiability result shows that identical visible observations may correspond to different egress conditions, so image-only observation is insufficient. The framework couples computer vision for visible demand, a spatio-temporal graph neural network that reconstructs demand at unobserved nodes by inductive node kriging, and a deterministic max-flow/min-cut solver that returns static network-level egress-capacity sufficiency and the binding cut. Because simultaneous competition for shared cut capacity admits more than one globally optimal allocation, per-zone sufficiency is reported as a solver-independent interval rather than a single value. Off-frame completion is evaluated externally on three real road-traffic benchmarks, where learning gives a consistent advantage under whole-node missingness but is outperformed by temporal interpolation under block missingness. The same methodology is then instantiated on a constructed 20-zone case-study graph of the Prophet's Mosque with simulated demand and a controlled perception-error proxy. In that controlled case study the completion-plus-analytic configuration yields lower egress-sufficiency error and more accurate bottleneck estimates than the incomplete configurations across the tested observability levels. No registered mosque imagery or measured zone-level demand was evaluated, so the study establishes the reasoning mechanism rather than deployment performance.

1. INTRODUCTION

Most deployed vision-based crowd-monitoring systems assume that crowd safety can be inferred from observable image features, with local density as the principal indicator of risk. That assumption is inconsistent with evidence from major crowd disasters [1]. Forensic analysis of the 2006 Mina/Jamarat disaster concluded that local density alone could not reliably identify when or where conditions became critical, and that crowd pressure gave substantially earlier warning [2]. The 2022 Itaewon crowd crush likewise arose where converging streams met an alley whose capacity could not dissipate the accumulated demand [3]. Pedestrian velocity has also been shown to represent perceived congestion more faithfully than local density [4].

The quantity governing crowd safety is therefore not local density but the relationship between pedestrian demand and the capacity of the surrounding network. That relationship is architectural: demand propagates through corridors, gates and open spaces, many of which lie outside the field of view of any single camera. A density that is acceptable in an open area may become hazardous when downstream capacity is limited, so systems relying only on image-derived density cannot recover the structural conditions that determine egress.

A second practical limitation arises from the sensing process itself: no camera observes an entire venue. Camera coverage is inherently fragmented, constrained by installation cost, placement, and field-of-view limitations. Consequently, crowd demand is observed only through a partial observation mask, with occupancy measured over the visible portion of the environment while the architectural structure governing pedestrian movement extends beyond the sensor's field of view. The corridors, gates, and intersections that determine whether accumulated demand can be discharged are therefore frequently unobserved, making this a problem of inference under structured, non-random missingness rather than image interpretation.

The proposed framework reformulates crowd-safety assessment as an architecture-aware operational reasoning problem under partial observation. It integrates computer vision, graph-based demand completion, and analytic egress evaluation to estimate operational conditions from incomplete visual observations. The framework consists of three sequential stages.

Stage 1-Perception (Visible Demand). A computer-vision (CV) crowd-counting model estimates pedestrian demand over the observed nodes of the venue graph from image data. This stage provides demand estimates for the visible portion

of the environment, while demand outside the camera field of view remains unobserved.

Stage 2-Off-Frame Demand Completion. Because analytic egress evaluation requires demand estimates over the complete venue graph, a spatio-temporal graph neural network (ST-GNN) reconstructs demand at unobserved nodes through inductive node kriging. The model exploits both spatial connectivity and temporal correlations to infer demand in regions that are not directly observable, thereby producing a complete demand field for subsequent analysis.

Stage 3-Analytic Egress Evaluation. The completed demand field is evaluated with a deterministic max-flow/min-cut formulation to estimate per-zone egress-capacity sufficiency and identify the binding bottlenecks. Because these quantities are derived analytically from the venue graph, the assumed capacity model, and established pedestrian-flow relationships, they are independent of learned annotations or simulation-specific labels.

Together, these stages transform partially observed image measurements into architecture-aware estimates of static network-level egress-capacity sufficiency. More broadly, the framework is an instance of reasoning under structured partial observation over a capacitated graph, in which a learned component recovers a latent demand field and an exact combinatorial solver maps that field to the operational answer; crowd safety is the application context in which it is developed and evaluated.

Despite progress in computer vision, crowd modeling and evacuation analysis, no existing framework supports architecture-aware egress reasoning under partial observation. Crowd-physics studies establish that density alone is insufficient but do not provide an estimation framework from partially observed imagery. Network-flow approaches formulate evacuation rigorously but assume the venue graph and the demand distribution are fully known. Graph neural networks applied to evacuation modeling [5] learn to reproduce a specific pedestrian simulator, so they remain tied to simulator-generated labels rather than analytically derived network-flow quantities.

A further limitation is the absence of datasets that jointly represent crowd imagery and the capacitated architectural graph of the same venue: crowd-counting datasets carry no egress topology, and evacuation or floorplan datasets carry no imagery. This work therefore develops an evaluation testbed combining a capacitated venue graph with simulated pedestrian demand and structured partial-observation conditions.

The contributions of this work are threefold.

(1) A constructive non-identifiability result for image-only egress reasoning. Identical visible scenes may correspond to arbitrarily different egress conditions, establishing that image-only observation is insufficient and that architectural structure must be represented for the competition-sensitive component of egress sufficiency.

(2) An architecture-aware formulation of off-frame demand completion under partial observation, together with an evaluation methodology that identifies where learned inference is required and where deterministic analysis suffices. The evaluation delimits the regime in which learning helps: whole-node reconstruction of unobserved regions, rather than spatial interpolation or temporal gap filling.

(3) Deterministic static network-level egress-capacity reasoning, coupled to the completion stage and evaluated on a constructed architectural case study with simulated demand under controlled partial observation, together with the evaluation testbed required to study that setting.

Scope. The framework is formulated over a generic capacitated venue graph under partial observation. The perception stage is evaluated on real crowd imagery from a public benchmark; the completion stage is evaluated externally on real road-traffic sensor networks; and the integrated completion-and-egress reasoning is evaluated on a constructed case-study graph of the Prophet's Mosque with simulated zone-level demand and a controlled perception-error proxy. The crowd-counting network is not run inside that case study, and the case-study graph has not been validated against authoritative venue records. The paper therefore claims estimation of egress-capacity sufficiency under the stated assumptions, rather than calibrated estimates of absolute risk, crowd-crush probability, or validated deployment performance.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature and identifies the research gap. Section 3 formulates the crowd-safety problem and presents the non-identifiability analysis. Section 4 describes the proposed architecture-aware framework. Section 5 outlines the experimental design and evaluation protocol. Section 6 presents the experimental results and discussion. Section 7 discusses threats to validity, Section 8 addresses reproducibility, and Section 9 concludes the paper.

2. RELATED WORK

The proposed framework combines concepts from five research areas: image-based crowd counting, spatio-temporal graph completion, network-flow evacuation modeling, crowd physics, and architecture-aware crowd reasoning. The following subsections review the most relevant work in each area and identify the specific gap addressed by this study.

2.1 Crowd counting and density estimation

The perception component of the proposed framework builds upon advances in image-based crowd counting, whose objective is to estimate pedestrian density or count directly from image data. Modern approaches are predominantly based on density-map regression, in which a convolutional neural network predicts a per-pixel density surface whose integral corresponds to the crowd count. Early multi-column architectures, such as MCNN [6], addressed scale variation using parallel convolutional branches with different receptive fields. CSRNet [7] subsequently simplified this design by combining a VGG-16 front-end with dilated convolutions, enabling larger receptive fields while preserving spatial resolution, and remains one of the most widely adopted crowd-counting models. Subsequent advances have further improved counting accuracy through Bayesian loss formulations [8], distribution-matching objectives (DM-Count) [9], and transformer-based architectures, while benchmark datasets have expanded from the ShanghaiTech datasets [6] to increasingly challenging high-density collections, including UCF-QNRF [10], JHU-CROWD++ [11], and NWPU-Crowd [12].

Two characteristics of this literature are particularly relevant to the present study. First, counting error is not uniform but increases systematically with crowd density, occlusion, perspective distortion, and other challenging

imaging conditions. This behavior is reflected in uncertainty-estimation methods [13], which demonstrate that both aleatoric and epistemic uncertainty correlate with prediction error. Second, existing crowd-counting methods operate exclusively on the observable image domain and therefore provide no information about architectural connectivity or downstream egress capacity beyond the camera field of view. Accordingly, CSRNet is employed only to estimate visible-zone pedestrian demand, while architecture-aware reasoning is introduced through graph-based demand completion and deterministic egress analysis.

2.2 Spatio-temporal graph kriging and imputation

Recovering signals at locations or time points for which no observations are available is commonly formulated as a graph kriging or imputation problem. Classical graph signal processing (GSP) approaches reconstruct missing values by exploiting signal smoothness over the graph through Laplacian regularization or harmonic interpolation [14, 15]. Related approaches, including temporal regularized matrix factorization (TRMF) [16], further exploit temporal dependencies to recover missing observations.

More recent methods employ graph neural networks to address the inductive setting, where missing nodes may not have been observed during training. IGNNK [17] performs inductive spatio-temporal kriging by learning graph diffusion operators over randomly masked subgraphs, enabling generalization to previously unseen sensors. GRIN [18] combines graph-based message passing with bidirectional recurrent architectures for multivariate time-series imputation, while subsequent studies have extended these approaches through uncertainty-aware graph imputation models.

These developments are directly relevant to the present work because off-frame crowd-demand estimation is naturally formulated as an inductive node-kriging problem, in which demand must be inferred for spatial locations that are entirely outside the camera field of view. Existing graph-kriging methods therefore provide the methodological foundation for the proposed demand-completion stage. The comparative performance of learned and classical interpolation methods under different missing-data regimes is evaluated experimentally in Section 6.

2.3 Network-flow models of building egress

Building evacuation has long been formulated as a capacitated network-flow problem [19, 20], in which rooms, corridors, and exits are represented as nodes and edges, while edge capacities are determined from architectural characteristics such as clear width and pedestrian-specific flow. Within this formulation, the maximum instantaneous evacuation rate from an occupied region to the exterior is equal to the value of the corresponding minimum cut. By flow-cut duality, the edges defining the minimum cut identify the architectural bottlenecks that constrain pedestrian movement. Dynamic extensions based on time-expanded networks, including quickest-flow and earliest-arrival formulations [21], further incorporate travel times into the evacuation process.

A common assumption underlying this literature is that both the venue graph and the associated demand distribution are fully specified before optimization. Consequently, existing network-flow methods address evacuation analysis given complete information, but they do not consider the complementary problem of recovering pedestrian demand under partial observation. The proposed framework addresses this limitation by combining graph-based demand completion with a deterministic network-flow formulation for operational egress analysis.

2.4 Crowd dynamics and safety

The limitations of density-only crowd-safety assessment are well established. The level-of-service framework [22] and the fundamental diagram describe the relationship between pedestrian density, speed, and specific flow, identifying the transition from stable movement to reduced flow as density increases. Forensic analysis of the 2006 Mina/Jamarat disaster demonstrated that local density alone could not reliably identify when or where conditions became critical [2]. Instead, crowd pressure, which combines pedestrian density and speed variance, provided a substantially more informative indicator of impending failure [2]. Pedestrian velocity has similarly been reported to reflect perceived congestion more accurately than conventional local-density measures [4]. Recent crowd disasters, including the 2022 Itaewon crush [3], further illustrate that hazardous conditions arise from the interaction between pedestrian demand and architectural bottlenecks rather than from local density alone.

Collectively, these studies motivate the use of architecture-aware egress sufficiency as an operational quantity for crowd-safety assessment and provide the physical basis for the capacity formulation adopted in this work through Fruin's specific-flow relationships. At the same time, they distinguish network-level capacity analysis from the higher-order crowd-pressure dynamics associated with sub-capacity turbulence, thereby clarifying the scope of the proposed framework.

2.5 Architecture-aware crowd reasoning

A growing body of work combines architectural representations with learning-based evacuation analysis. The closest study to the present work [5], which trained an edge-featured graph neural network on BIM-derived building graphs to predict per-node density classes generated by a pedestrian simulation engine and evaluated the model against labels produced by the same simulator. Related approaches, such as DRAGON [23], address evacuation planning or routing once a risk or congestion field has already been estimated.

Two common limitations characterize this literature. First, the learning objective is typically defined using simulator-generated labels and evaluated within the same simulation environment, restricting assessment to simulator-consistent predictions. Second, these approaches assume that pedestrian demand is available throughout the complete architectural graph and therefore do not address estimation under partial observation. Scene-graph reasoning [24] provides a related methodology for modeling object relationships within images; however, it is fundamentally limited to information contained within the observable scene and is therefore not designed to recover operational information that lies outside the camera field of view.

Consequently, existing architecture-aware learning methods do not address crowd-safety estimation from partially observed image data, nor do they consider the non-identifiability introduced by incomplete architectural observation, whereby identical visible scenes may correspond

to substantially different operational egress conditions.

2.6 Synthesis and research positioning

Across this literature a common limitation recurs. Crowd-counting methods estimate visible demand without architectural reasoning; graph-kriging approaches reconstruct missing signals but are not formulated for image-driven egress estimation; network-flow methods analyze egress rigorously but assume complete knowledge of demand; crowd-dynamics studies establish that hazard arises from the interaction of demand and infrastructure without providing an estimation framework; and architecture-aware learning methods assume complete graph observation and simulator-derived supervision. The framework proposed here addresses that gap directly. Table 1 summarizes the capabilities of each methodological family.

Table 1. Comparison of representative methodological families

Method Family	Visible Demand	Graph	Architecture-Aware Egress	Partial Obs.
Crowd counting (MCNN, CSRNet)	√	×	×	×
Network-flow egress (max-flow/min-cut)	×	√	√	×
ST graph kriging (IGNNK, GRIN)	×	√	×	√
Architecture-aware evacuation (BIM-GNN)	×	√	√	×
Proposed Framework	√	√	√	√

Note: √ denotes native support; × denotes a capability assumed away or absent. Visible Demand: recovers demand from imagery; Graph: reasons over a relational graph; Architecture-Aware Egress: reasons over never-observed architectural capacity; Partial Obs.: operates with entirely unobserved nodes. Among the representative methodological families considered here, the proposed framework combines image-based demand estimation, graph-based reasoning, architecture-aware egress analysis, and operation under partial observation within a unified framework.

3. PROBLEM FORMULATION

3.1 Capacitated venue graph

A venue is represented as a directed capacitated venue graph,

$$G = (V, E, c, \tau), \tag{1}$$

constructed offline from architectural information and floorplan geometry. In this work the framework is instantiated in-domain on the Prophet's Mosque (Al-Masjid an-Nabawi), Madinah, which serves as the case study throughout. The node set V represents the functional regions of the venue, including gates, courtyards, halls, galleries, corridors, ramps, designated bottlenecks, and a virtual super-sink t representing safe egress from the venue. The super-sink is implemented as an IndoorGML Node-Relation-Graph anchor node [25]. The edge set $E \subseteq V \times V$ encodes directed connectivity between adjacent regions. Directed edges are retained to represent asymmetric pedestrian movement through one-way passages, ramps, and controlled egress routes.

Each edge $e \in E$ is associated with two attributes:

• a capacity

$$c(e) = w(e) f_s, \tag{2}$$

expressed in persons·s^{-1}, where, $w(e)$ denotes the effective clear width of the edge (m) and f_s denotes the pedestrian-specific flow (persons·m^{-1}·s^{-1}); and

• a traversal time

$$\tau(e) = \ell(e)/v_{\text{free}}, \tag{3}$$

where, $\ell(e)$ is the edge length and v_{free} is the free-flow walking speed. This quantity is required for dynamic extensions based on quickest-flow formulations.

Throughout this work, edge capacities are determined from architectural geometry through Eq. (2), following established pedestrian-flow models [2, 22]. The specific-flow parameter fs is fixed at the Fruin free-flow value for all reported results. Because every edge capacity scales linearly with fs, a global change in fs rescales all cut capacities identically and leaves the identity of the binding cut unchanged. The sufficiency endpoints themselves scale with fs and are therefore conditional on the assumed specific-flow value. Clear widths $w(e)$ are recorded with per-edge provenance flags and the majority are estimated rather than surveyed, so the resulting capacities are conditional assumptions of the case-study model rather than measured quantities.

3.2 Observation model

Let Π denote the camera projection that maps a physical scene to an image,

$$I = \Pi(\text{scene})$$

The projection provides only a partial observation of the physical environment by restricting the observable information to the camera field of view. Consequently, the node set is partitioned into an observed subgraph and its complement,

$$V_{obs} \subseteq V, V \setminus V_{obs}$$

where, $V \setminus V_{obs}$ denotes the off-frame nodes. A binary observation mask,

$$M: V \to \{0,1\} \tag{4}$$

records whether each node is visible at a given instant. The image therefore contains information only about the induced subgraph,

$$G_{\text{obs}} = G[V_{\text{obs}}]$$

while no visual information is available for nodes outside the camera field of view.

A key characteristic of the proposed formulation is the asymmetry between the venue structure and the pedestrian demand. The architectural graph G, including its topology and

edge capacities, is assumed to be known from the offline venue model, whereas pedestrian demand on the off-frame nodes remains unobserved during deployment. Consequently, only the demand field must be inferred, while the underlying architectural structure is fully specified.

3.3 Demand and egress sufficiency

Let $d: V \to \mathbb{R}_{\geq 0}$ denote the pedestrian demand over the venue graph, where d(v) is the per-node required egress rate (persons·s−1): the rate at which pedestrians at node v must be discharged toward the exits at the evaluated instant, rather than a static occupancy count. Because the perception stage measures an occupancy N(v) in persons while edge capacities are expressed in persons·s−1, the two are related through a declared clearance horizon T_{evac}: $d(v) = N(v)/T_{evac}$. T_{evac} is a modelling scenario parameter, not an empirically established safety threshold for any particular venue; the horizon used in the case study and the sensitivity range over which the conclusions were checked. For a set of nodes $A \subseteq V$ the summation convention $d(A) = \sum_{v \in A} d(v)$ is adopted, and the same convention applies to the estimated and completed demand fields $\hat{d}$ and $\tilde{d}$. The CV front-end provides a point estimate $\hat{d}(v)$ only for the observed nodes V_{obs}. Consequently, demand over the unobserved nodes $V \setminus V_{obs}$ must be inferred before venue-wide egress analysis can be performed.

For a source region S_z corresponding to zone z, the maximum instantaneous egress rate is given by the value of the minimum cut separating S_z from the sink node t. The corresponding egress-capacity sufficiency is defined as,

$$\rho(z) = \frac{\text{MinCut}(S_z \to t; c)}{\sum_{v \in S_z} d(v)}, \tag{5}$$

where, the numerator represents the throughput capacity of the corresponding minimum cut and the denominator the total egress rate that must be discharged through zone z; both carry units of persons·s^{-1}, so the ratio is dimensionless. Eq. (5) is evaluated for zone z on its own, so the bounded form $s_{max}(z) = min\{1, \rho(z)\} \in [0, 1]$ is the ceiling clearable fraction available to z when it does not contend for shared capacity. When all zones are loaded simultaneously the total cleared flow is the global maximum-flow value, but the way that flow is divided among zones is in general not unique: several globally optimal allocations exist. Both endpoints below are independent of which maximum flow a solver happens to return; the ceiling endpoint $s_{max}(z)$ is obtained from Eq. (5). Writing F* for the set of flows attaining that global maximum value from the super-source S over all loaded zones to t, the guaranteed endpoint is,

$$s_{min}(z) = \min_{f \in F_*} \frac{f(S \to z)}{d(S_z)} \tag{6}$$

Every globally optimal allocation therefore satisfies $s_{min}(z) \leq s(z) \leq s_{max}(z)$, and $\Delta_s(z) = s_{max}(z) - s_{min}(z)$ measures the allocation ambiguity induced by competition for shared cut capacity. The two endpoints are solver-independent: $s_{max}(z)$ follows from Eq. (5) and $s_{min}(z)$ is obtained by linear programming over F*. A value of $s_{max}(z) = 1$ denotes a zone whose offered demand is fully clearable when it is not contending, and $s_{min}(z) = 1$ a zone that clears fully under every optimal allocation.

The egress-sufficiency interval provides the basis for two operational outputs:

(1) a zone-level egress-capacity sufficiency interval [$s_{min}(z)$, $s_{max}(z)$]; and

(2) identification of the binding cut-edge set C* of the loaded network.

The formulation estimates relative egress-capacity sufficiency under the stated modeling assumptions. It is a static network-capacity model: it does not represent pedestrian interactions, counterflows, queue formation or spillback, congestion propagation, crowd turbulence, behavioral response, dynamic route choice, or time-dependent evacuation, and it does not estimate absolute risk, crowd-crush probability, or evacuation time. Crowd safety is the application context; egress-capacity sufficiency is the computed quantity.

3.4 Partial-observation demand completion

Evaluation of the egress-sufficiency ratio in Eq. (5) requires pedestrian demand over every node contributing to the corresponding minimum cut, including nodes outside the camera field of view. Consequently, the deployment problem is not the solution of a fully specified network-flow instance, but the recovery of the complete demand field under partial observation. Formally, the objective is,

$$\begin{gathered} \text{given } (G, \hat{d}|_{V_{obs}}, \sigma_d|_{V_{obs}}, M), \\ \text{recover } \tilde{d} \text{ on } V \setminus V_{obs}, \\ \text{then compute } \rho(z) \text{ using } \tilde{d}. \end{gathered} \tag{7}$$

Because $MinCut(Sz \to t)$ depends on architectural elements that may lie outside the observable region, identical visible observations may correspond to different egress conditions. Consequently, no image-only predictor $f(I)$ can correctly estimate the egress condition for every matched-density, mismatched-architecture pair. Recovering the missing demand over the known architectural graph therefore becomes the central inference problem addressed in this work. Once the completed demand field $\tilde{d}$ has been obtained, egress sufficiency is evaluated deterministically using the max-flow/min-cut formulation introduced in the previous subsection.

Throughout this work, the proposed completion model is evaluated only for the whole-node (off-frame) reconstruction setting. Short within-node temporal interpolation is treated as a separate problem and evaluated independently.

3.5 Non-identifiability under partial observation

The following proposition formalizes a fundamental property of operational egress estimation under partial observation.

Proposition 1 (Non-identifiability under Partial Observation). There exist two venue graphs that produce identical observable images while yielding different minimum cuts. Consequently, operational egress sufficiency is not identifiable from image observations alone.

Proof (by construction). Consider two venue graphs, G_A and G_B, that share the same visible subgraph and therefore produce an identical image observation for a visible node x. In graph G_A,

$$x \to \{u_1, u_2, u_3\}$$

three independent downstream exits are available. In graph G_B,

$$x \to y \to u_1$$

where, the intermediate node *y* lies outside the camera field of view and u_1 represents a restricted-capacity exit node. Because the downstream architectural structure is unobserved, both graphs generate the same image despite having different connectivity.

Consequently,

$$\mathrm{MinCut}_{G_A}(S_z \to t) \neq \mathrm{MinCut}_{G_B}(S_z \to t)$$

which implies,

$$\rho_{G_A}(z) \neq \rho_{G_B}(z)$$

Thus, identical image observations may correspond to different operational egress conditions. Therefore, no estimator based solely on the image observation, $f(I)$ can uniquely recover operational egress sufficiency for all matched-density, mismatched-architecture pairs.

Proposition 1 motivates the formulation adopted in this work. Rather than estimating operational egress directly from the observable image, the proposed framework first reconstructs pedestrian demand over the known architectural graph and subsequently evaluates egress sufficiency using the deterministic max-flow/min-cut formulation.

4. METHODOLOGY

4.1 Visible-demand estimation

The perception stage estimates pedestrian demand over the observable portion of the venue graph. Each registered image region is mapped to a per-node demand estimate,

$$\hat{d}(v), v \in V_{\mathrm{obs}}$$

which provides the visible component of the demand field.

CSRNet [7] is adopted as the perception front-end because it is a well-established density-map regression model with strong benchmark performance. The network combines a VGG-16 feature extractor with dilated convolutions to predict a crowd-density map whose spatial integral yields the estimated pedestrian count. The proposed framework is independent of the underlying crowd-counting architecture, and any density-estimation model capable of producing per-region pedestrian counts may be substituted.

4.2 Architectural graph construction

The architectural graph is constructed offline from publicly available descriptive information about the venue, including published area and capacity figures together with publicly available plan and imagery sources, and is organized following the IndoorGML Node-Relation Graph (NRG) model [25]. It is a constructed case-study representation rather than a surveyed or officially validated model of the venue. The graph provides a topology-preserving representation in which functional spaces are nodes and pedestrian connectivity is given by directed edges.

Directed edges encode pedestrian passability between adjacent regions. Edge capacities are computed according to Eq. (2), in which $w(e)$ denotes the effective clear width and fs denotes the pedestrian-specific flow. Published dimensions are used where available; the remaining widths are estimated from public plan and imagery sources and are flagged as inferred in the released configuration. Those width intervals are recorded as provenance metadata and are not propagated through the reported analysis, so the capacities used here should be read as point assumptions.

Each node also carries the binary observability indicator of Eq. (4), which partitions the demand field into observed and unobserved components and provides the input to the completion stage.

4.3 Spatio-temporal demand completion

The second stage of the proposed framework reconstructs pedestrian demand over the unobserved portion of the venue graph. Given the observed demand estimates,

$$\hat{d}(v), v \in V_{\mathrm{obs}}$$

together with the architectural graph *G*, the objective is to estimate pedestrian demand over the off-frame nodes, yielding a completed field $\tilde{d}$ defined on all of *V* that coincides with $\hat{d}$ on V_{obs},

$$\tilde{d}(v), v \in V \setminus V_{\mathrm{obs}}$$

Demand completion is formulated as an inductive node-kriging problem and implemented using an IGNNK-style spatio-temporal graph neural network [17]. The model operates over a sliding temporal window of observed demand and exploits both the spatial connectivity of the architectural graph and the temporal evolution of pedestrian demand to reconstruct the missing node values.

An inductive architecture is adopted because the deployment graph is not assumed to match the graphs seen during training. Because the architecture uses no node-specific learned parameters, its parameterization is compatible with different graph sizes and topologies; cross-graph inference without retraining is not evaluated here, and for the case-study graph the completion model is refitted from scratch.

The proposed framework employs the graph neural network exclusively for whole-node (off-frame) demand completion. Once the complete demand field has been reconstructed, the resulting estimates are passed to the analytic egress solver described in the following subsection.

The proposed completion model is evaluated against representative classical and learning-based baselines.

4.4 Analytic egress solver

Once the complete demand field $\tilde{d}$ has been reconstructed, operational egress sufficiency is evaluated using an analytic network-flow formulation. A maximum-flow algorithm (the preflow-push implementation of NetworkX in the reported experiments) is applied to compute the corresponding minimum cut,

$$\mathrm{MinCut}(S_z \to t)$$

from which the binding cut-edge set C* is obtained directly from the residual graph. The ceiling endpoint of Eq. (5) is then

evaluated on the completed demand field as,

$$\rho(z) = \frac{\text{MinCut}(S_z \to t; c)}{\sum_{v \in S_z} \tilde{d}(v)} \tag{8}$$

Per-zone egress sufficiency is reported as the interval $[s_{min}(z), s_{max}(z)]$. Loading all zones simultaneously fixes the total cleared flow but not its division among zones, so a single realized allocation would depend on which optimal flow the solver returns; the two endpoints are computed instead, $s_{max}(z)$ from Eq. (5) and $s_{min}(z)$ by minimizing the flow delivered to zone z over all globally optimal flows. The interval width $\Delta_s(z)$ is the range of per-zone outcomes attributable to competition for shared cut capacity.

The resulting solution provides three operational outputs:

(1) a zone-level egress-sufficiency interval $[s_{min}(z), s_{max}(z)]$ for each venue zone;

(2) the binding cut-edge set C*; and

(3) a zone-level map of the egress-sufficiency intervals.

Edge traversal times $\tau(e)$ are retained in the graph representation for future time-expanded quickest-flow extensions; they are not used by the static formulation evaluated in this paper.

The proposed framework adopts an analytic rather than a learned egress solver. This design confines learning to the demand-completion stage while exploiting the exact solution provided by the max-flow/min-cut formulation once the complete demand field has been estimated.

4.5 Proposed inference workflow

Figure 1 summarizes the proposed inference workflow. The architectural graph $G = (V, E, c, \tau)$ is constructed offline and remains fixed during deployment. At inference time, the CV front-end estimates visible pedestrian demand, producing per-node demand estimates $\hat{d}$ over the observed node set V_{obs} and the corresponding observation mask M.

The spatio-temporal graph neural network then reconstructs demand over the unobserved nodes $V \setminus V_{obs}$, yielding the completed field $\tilde{d}$, which is supplied to the analytic max-flow/min-cut solver. The solver returns the per-zone egress-capacity sufficiency interval $[s_{min}(z), s_{max}(z)]$ and the binding cut-edge set C*.

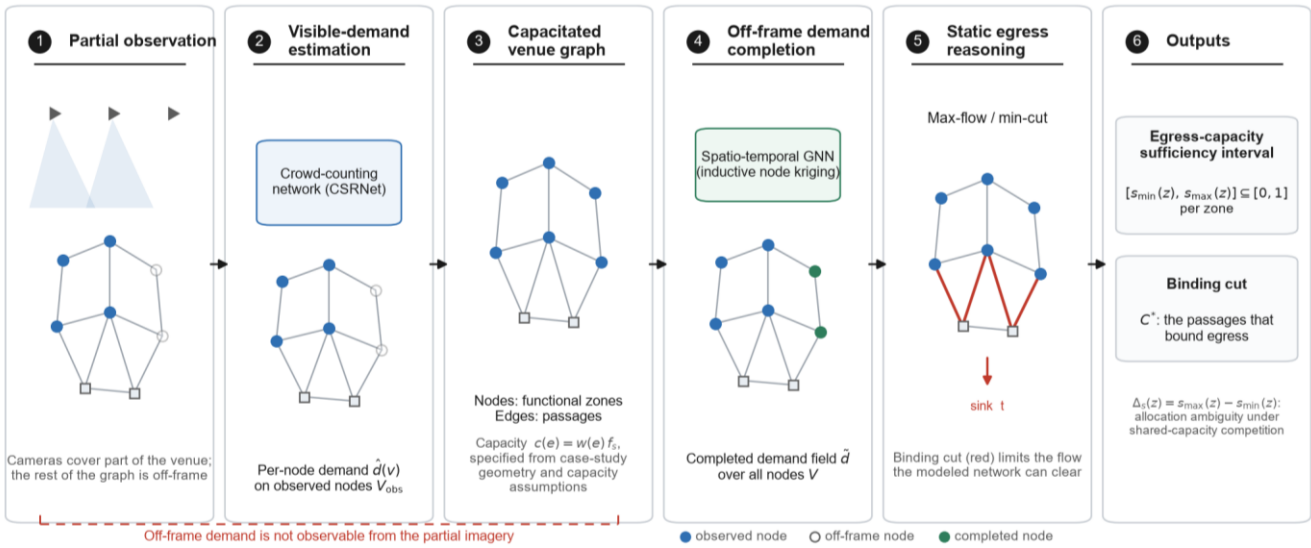


Figure 1. The proposed architecture-aware framework

Note: A fixed architectural venue graph supplies topology and capacities; visible demand is estimated from camera observations, completed over off-frame nodes by a spatio-temporal GNN, and analyzed by a max-flow/min-cut solver to produce per-zone egress-capacity sufficiency intervals and the binding cut.

5. EXPERIMENTAL DESIGN

The framework is evaluated through five experiments that isolate its components, reported one-to-one as R1–R5. The evidence they provide is of four distinct kinds: real crowd imagery (E1), controlled synthetic constructions (E2), real graph-structured traffic measurements (E3, E4), and a constructed architectural case study with simulated demand (E5). This section describes the protocol, datasets, metrics and decision criteria for each.

5.1 E1-visible-demand estimation

Experiment E1 characterizes the accuracy of the CV demand estimator and its dependence on crowd density, establishing the reliability of the pipeline input.

CSRNet is evaluated on the complete ShanghaiTech Part B test set, comprising 316 real-world crowd images with ground-truth counts ranging from 9 to 539 pedestrians. Overall performance is quantified using the mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE) and mean estimation bias.

To examine the effect of crowd density, the test images are partitioned into four ground-truth count intervals: [0, 50), [50, 100), [100, 200), and [200, ∞). For each interval, MAE, RMSE, mean bias, and MAPE are reported.

The resulting density-dependent error profile is reported as a measured characterization of perception error. It is not used to parameterize the perception-error perturbation applied in Experiment E5, and no calibrated uncertainty is propagated through the subsequent stages.

5.2 E2-controlled non-identifiability demonstration

Experiment E2 evaluates the practical implications of Proposition 1 using paired synthetic scenes under two regimes.

In the easy regime, paired scenes contain observable differences that make the operational ordering comparatively recoverable. In the topology-hard regime, visible pedestrian demand is matched while downstream architectural topology is varied, so paired scenes contain identical or near-identical observable demand information but different egress capacities.

To attribute performance gains to specific sources of information, a hierarchical baseline ladder is constructed:

•**B0-Vision-only CNN**: image observations only.

•**B1-Vision + architectural descriptors**: image observations augmented with hand-engineered architectural features (e.g., exit count, minimum exit width, and openness).

•**B2-Architecture-only GNN**: architectural graph without pedestrian demand.

•**B3-Proposed graph model**: architectural graph with per-node pedestrian demand.

•$\boldsymbol{B}_{\text{analytic}}$**-Analytic impute-then-solve**: reconstructed demand followed by exact max-flow/min-cut computation.

•**Oracle**: exact max-flow/min-cut using the ground-truth demand field.

Performance is evaluated using pair-ordering accuracy and the Oracle gap. The easy regime is noiseless and the topology-hard regime applies a controlled multiplicative perception-error perturbation of relative magnitude 0.15; no sweep over perturbation levels is reported. Together, these metrics quantify the ability of each baseline to distinguish operationally different matched pairs despite identical observable imagery.

5.3 E3-spatial demand completion

Experiment E3 evaluates whether a learned spatial graph model provides an advantage over established non-learned interpolation methods for reconstructing demand from a single graph snapshot. The experiment isolates the spatial component of the completion problem by excluding temporal information.

An IGNNK-style spatial graph neural network is compared with representative classical baselines, including global mean imputation, graph-based k-nearest neighbors (kNN-on-graph), inverse-distance weighting, harmonic interpolation using the graph Laplacian (Laplacian-GSP), and low-rank singular value decomposition (SVD). Evaluation is performed on two complementary settings: (i) the real PeMSD7-M traffic network comprising 228 Caltrans sensors with the corresponding road-network topology, and (ii) a real-topology, simulated-demand setting used to assess generality.

Performance is measured using MAE and the coefficient of determination (R^2) on masked nodes under masking ratios of 25%, 50%, and 75%, averaged across multiple random masking realizations.

5.4 E4-spatio-temporal demand completion

Experiment E4 evaluates the only learned component of the proposed framework: spatio-temporal demand completion under two complementary missingness regimes, whole-node (off-frame) missingness and temporal-block missingness. The objective is to determine whether temporal graph learning provides an advantage over representative classical interpolation methods in the deployment scenario for which the framework is designed.

The evaluation is performed on three canonical traffic datasets-PEMS04, PEMS08, and PeMSD7-M-which provide real sensor measurements together with the corresponding graph topology. An IGNNK-style inductive spatio-temporal graph neural network is compared with a comprehensive set of classical baselines, including temporal mean, persistence, temporal linear interpolation, low-rank SVD, TRMF, graph-based *k*-nearest neighbors, Laplacian graph signal processing (Laplacian-GSP), and spatio-temporal graph signal processing (ST-GSP).

To distinguish spatial and temporal interpolation regimes, the evaluation considers two complementary missing-data settings. In the node-missingness setting, complete sensor nodes are withheld throughout the evaluation interval, representing permanently unobserved (off-frame) regions. In the block-missingness setting, contiguous temporal segments are removed from otherwise observed nodes, representing conventional temporal interpolation.

Performance is evaluated over all combinations of the three datasets, three masking ratios (25%, 50%, and 75%), and the two missing-data regimes. MAE is computed across multiple random masking realizations and independent training seeds.

Significance is assessed with seed-aligned paired permutation tests and Benjamini–Hochberg correction across each family of comparisons. Case-study runs are analyzed separately because repeated runs vary only the masking, training and perturbation realizations.

5.5 E5-integrated architecture-aware case-study evaluation

Experiment E5 evaluates the architecture-aware completion and egress-reasoning stages composed on the constructed 20-zone case-study graph of the Prophet's Mosque. Its purpose is to test whether those stages operate correctly together on a realistic venue topology under controlled partial observation. It is not an image-to-safety deployment evaluation.

The architectural graph, edge capacities, spatio-temporal graph neural network, analytic max-flow/min-cut formulation, and evaluation protocol remain unchanged. Two inputs, however, cannot be taken from measurement. First, no registered Prophet's Mosque crowd imagery was available, so the crowd-counting network is not run in this experiment; the perception stage is represented by a controlled multiplicative perception-error proxy of relative magnitude 0.10 applied to the visible zones. The proxy is unbiased and is not derived from Experiment E1, whereas the measured counter undercounts systematically with density, so the perception input used here is optimistic. It perturbs the visible observations only, while the completion model is fitted and scored on the underlying demand series, so the perception-to-completion interface is not exercised. Second, no public dataset provides measured zone-level pedestrian occupancy for the Prophet's Mosque, so per-zone occupancy is simulated with a graph-coupled prayer-cycle model whose spatial coupling derives from the same capacity-weighted adjacency later used by the completion model. The simulated occupancy is converted to the required egress rate through the declared horizon, $d(v) = N(v)/T_{evac}$, with a reference scenario of $T_{evac} = 900$ *s*. That value is a presentation scenario fixed before the results reported below were examined; it is not a standard and is not a recommended clearance time for the venue. All conclusions reported below were checked over the predeclared range $T_{evac} \in \{120, 300, 600, 900, 1200, 1500, 1800, 2400, 3000\}$ s, which spans network loadings from $\rho \approx 12.9$ to $\rho \approx 0.5$. Because the crowd-counting network is not run in this

experiment, Eq. (5) and the horizon define the intended dimensional interface, but the complete perception-to-egress interface is not empirically validated here.

The demand series is split chronologically, the completion model being fitted on the earlier portion and evaluated on the held-out remainder. The off-frame zones are therefore historically observed during fitting but unobserved throughout evaluation, whereas the non-learned prior uses only the concurrent visible values.

Rather than the six-rung ladder of Experiment E2, the in-domain evaluation compares three deployment configurations of increasing capability-CV-only, CV + analytic, and the CV + ST-GNN + analytic configuration-which isolate the contributions of perception, analytic egress reasoning, and off-frame demand completion respectively.

6. RESULTS AND DISCUSSION

Results are reported one-to-one against the experiments: R1 (6.1) the perception stage, R2 (6.2) the controlled non-identifiability demonstration, R3 (6.3) spatial completion, R4 (6.4) spatio-temporal completion on real traffic networks, and R5 (6.5) the integrated case study. Quantitative results are reported as the mean over repeated runs with the stated dispersion.

6.1 R1-visible-demand estimation

On the ShanghaiTech Part B test set, CSRNet attains an MAE of 23.18, an RMSE of 36.22 and a MAPE of 18.0%. Table 2 stratifies this by density: MAE rises from 5.38 pedestrians in the sparsest interval to 58.82 in the densest, and the mean bias falls from −0.81 to −58.71, so the counter systematically undercounts as density increases. This density-dependent bias is the measured error characteristic of the perception stage.

6.2 R2-controlled non-identifiability demonstration

Experiment E2 is a controlled demonstration of Proposition 1 rather than independent evidence of real-world performance. The easy regime contains observable differences, whereas the topology-hard regime is constructed so that visible pedestrian demand is matched while the discriminating architectural information lies outside the observable field of view. Within this controlled construction, the experiment evaluates whether operational egress conditions can be inferred from image observations alone under partial observation. Table 3 summarizes the performance of the baseline ladder across the easy and topology-hard evaluation regimes.

Table 2. CSRNet counting error by density bin, ShanghaiTech Part B test split

GT Count Bin	n	GT Mean	Pred Mean	MAE	RMSE	Mean Bias	MAPE (%)
[0, 50)	60	35.57	34.76	5.38	6.78	−0.81	17.0
[50, 100)	109	70.87	64.39	12.10	24.98	−6.48	16.9
[100, 200)	97	149.32	124.54	28.29	32.45	−24.78	19.1
[200, ∞)	50	295.56	236.85	58.82	69.52	−58.71	19.3
Overall	316	123.80	104.52	23.18	36.22	−19.28	18.0

Note: mean absolute error (MAE), root mean squared error (RMSE), root mean squared error (RMSE).

Table 3. Controlled matched-pair demonstration of Proposition 1

Baseline	Easy		Hard (σ = 0.15)	
	Pair-Order	Oracle Gap	Pair-Order	Oracle Gap
B0 (Vision only)	0.000	1.000	0.375	0.625
B1 (Vision + architectural descriptors)	0.917	0.083	0.375	0.625
B2 (Architecture graph)	0.917	0.083	0.694	0.306
B3 (Graph + demand)	0.917	0.083	0.431	0.569
$B_{analytic}$	1.000	0.000	1.000	0.000
Oracle	1.000	0.000	1.000	0.000

Note: The easy regime contains topology-distinguishable scene pairs, whereas the hard regime contains matched-density, mismatched-architecture pairs under multiplicative perception noise of relative magnitude 0.15. Pair-ordering accuracy measures the ability to recover the correct operational ordering, and the Oracle gap denotes the performance difference relative to the exact analytic solution.

In the easy regime, all baselines except the vision-only model achieve high pair-ordering accuracy, indicating that operationally distinct scenes can be separated when the relevant architectural differences remain directly observable or readily recoverable. By contrast, the vision-only baseline (B0) orders the pairs at 0.000, that is, systematically inverted relative to the ground-truth ordering and below the chance level of 0.5, reflecting that higher visible density is anti-correlated with egress capacity in these pairs. Visible density therefore carries no directly usable operational ordering even under comparatively favorable conditions.

The topology-hard regime exposes the limitation predicted by Proposition 1. When scene pairs share nearly identical visible pedestrian distributions but differ in their off-frame architectural topology, the vision-only baseline (B0) achieves a pair-ordering accuracy of only 0.375. Augmenting the image with hand-engineered architectural descriptors (B1) provides no measurable improvement, also yielding 0.375. These results indicate that neither visible crowd density nor simple architectural summary statistics contain sufficient information to recover operational egress conditions when the discriminating information resides outside the observable field of view.

Introducing explicit architectural topology improves discrimination. The architecture-only graph model (B2) increases pair-ordering accuracy from 0.375 to 0.694, indicating that the venue graph carries information unavailable from image observations alone. The accuracies in Table 3 are single point estimates over a finite pair set and are reported without uncertainty intervals, so only the larger differences

should be interpreted. Notably, supplying per-node demand to the graph model (B3) does not improve on B2 and in fact reduces pair-ordering accuracy to 0.431; this is reported as an open negative result rather than evidence that demand is uninformative, since the remaining gap to the Oracle is closed by the analytic solver below.

The final two baselines establish the appropriate division of labor between learning and analytic reasoning. The analytic impute-then-solve baseline (Banalytic) reproduces the Oracle in both regimes, producing a zero Oracle gap. That agreement is expected by construction, because both apply the same deterministic max-flow/min-cut formulation to a complete demand field; it is an exactness check on the solver rather than evidence of model superiority. Once a complete demand field is available, the static operational egress problem is therefore solved exactly by the analytic max-flow/min-cut formulation. Neither learned rung improves upon it, because the synthetic matched-pair experiment isolates the static egress solve rather than the demand-completion problem for which learning is intended. Rather than motivating a learned flow solver, this result reinforces the proposed decomposition of the framework: learning is required only to reconstruct off-frame pedestrian demand, whereas operational egress evaluation itself is most appropriately performed using the exact analytic max-flow/min-cut formulation.

6.3 R3-spatial demand completion

Experiment E3 isolates the spatial component of the completion problem by reconstructing demand from a single graph snapshot, with temporal information excluded. On the real PeMSD7-M sensor graph the learned spatial model does not improve on the classical interpolators: at 25% masking it attains an MAE of 7.60 ± 0.10 against 7.55 ± 0.08 for both inverse-distance weighting and Laplacian graph signal processing, and 5.78 ± 0.16 for low-rank matrix completion; it remains outside the leading group at 50% and 75% masking. On the geometric graph with simulated demand the learned and closed-form estimators differ by less than the spread across seeds at every masking ratio.

The spatial-only setting therefore provides no evidence that learning is required. This negative result delimits the claim made in this paper: the value of the learned component is specific to reconstruction that exploits temporal structure across entirely unobserved nodes, and not to spatial interpolation as such.

6.4 R4-spatio-temporal demand completion on real traffic networks

Experiment E4 evaluates demand completion across eighteen configurations: three datasets × three masking ratios × two missingness regimes. The regimes are reported separately because their outcomes differ.

Table 4 reports the numerical results for the whole-node (off-frame) regime and Table 5 the corresponding statistical significance. Across all nine node-missingness configurations (three datasets × three masking ratios), the proposed spatio-temporal graph model achieves the lowest reconstruction error, outperforming the strongest graph-based classical baseline in every such case. The improvement is largest on the sparse traffic-flow graphs (PEMS04 and PEMS08), where the reduction in MAE ranges from 33% to 62%, while smaller but consistent improvements of 13–15% are observed on the denser PeMSD7-M graph against the ST-GSP baseline. These margins should be read together with the information protocol: under node missingness the learned model is standardized using per-node training-period statistics, which are available for the withheld nodes, whereas the classical baselines receive only the test snapshots. The comparison therefore reflects a setting in which the hidden nodes were historically observed-the intended deployment condition-but it is not an equal-information comparison.

The advantage is regime-specific. In the complementary block-missingness setting, which represents short temporal gaps within otherwise observed nodes, temporal linear interpolation outperforms the graph neural network in all nine configurations, by 27.7% to 53.9% in MAE; Table 6 reports all nine configurations. Together with the spatial-only result, this indicates that learning provides a measurable advantage only when reconstructing demand for entirely unobserved graph nodes, whereas classical interpolation remains preferable for conventional temporal gap filling.

The statistical analysis confirms that the observed improvements are consistent across repeated training runs. Across the nine node-missingness evaluation configurations, the proposed model improves on the strongest classical baseline for every one of the five seeds, under a seed-aligned exact one-sided paired sign-flip permutation test. With five paired seeds the smallest attainable p-value is $2^{-5} = 0.031$, which every comparison attains; the identical values therefore indicate a uniform win across seeds at the resolution limit of the test rather than nine independent levels of evidence. Because the raw p-values are identical, Benjamini–Hochberg correction returns $q = p$ and adds no further discrimination. The test establishes consistency across training runs on fixed benchmark datasets rather than generalization across independent environments. Large paired effect sizes ($d_z = 4.8$ – 13.8) indicate that the improvements are substantial rather than merely statistically significant, while the Wilcoxon signed-rank test yields consistent conclusions.

Table 4. Whole-node completion on real sensor graphs: IGNNK against the strongest classical baseline

Dataset	Mask Frac.	Best Classical Baseline	Classical MAE	IGNNK MAE	Improvement (%)
PEMS04	25%	kNN-on-graph	74.46 ± 2.13	41.56 ± 2.83	44.2
PEMS04	50%	kNN-on-graph	83.59 ± 4.40	47.01 ± 2.54	43.8
PEMS04	75%	kNN-on-graph	88.13 ± 2.72	59.15 ± 4.08	32.9
PEMS08	25%	kNN-on-graph	77.16 ± 5.25	31.79 ± 3.46	58.8
PEMS08	50%	kNN-on-graph	88.10 ± 4.21	33.24 ± 1.66	62.3
PEMS08	75%	kNN-on-graph	84.81 ± 3.24	38.79 ± 2.24	54.3
PeMSD7-M	25%	ST-GSP	7.46 ± 0.23	6.34 ± 0.16	15.0
PeMSD7-M	50%	ST-GSP	7.35 ± 0.14	6.32 ± 0.19	14.0
PeMSD7-M	75%	ST-GSP	7.45 ± 0.12	6.44 ± 0.09	13.5

Note: mean absolute error (MAE); spatio-temporal graph signal processing (ST-GSP); k-nearest neighbors (kNN).

Table 5. Paired significance tests for the whole-node completion results

Dataset	Frac.	Best Classical Baseline	MAE Red. (%)	95% CI (MAE)	p_{perm}	q_{BH}	d_z
PEMS04	25%	kNN-on-graph	44.2	[29.83, 35.97]	0.031	0.031	8.21
PEMS04	50%	kNN-on-graph	43.8	[33.07, 40.04]	0.031	0.031	8.41
PEMS04	75%	kNN-on-graph	32.9	[24.79, 34.16]	0.031	0.031	4.79
PEMS08	25%	kNN-on-graph	58.8	[38.92, 52.93]	0.031	0.031	5.04
PEMS08	50%	kNN-on-graph	62.3	[50.90, 58.17]	0.031	0.031	11.63
PEMS08	75%	kNN-on-graph	54.3	[43.62, 48.94]	0.031	0.031	13.77
PeMSD7-M	25%	ST-GSP	15.0	[0.96, 1.28]	0.031	0.031	5.55
PeMSD7-M	50%	ST-GSP	14.0	[0.96, 1.09]	0.031	0.031	12.52
PeMSD7-M	75%	ST-GSP	13.5	[0.96, 1.09]	0.031	0.031	11.53

Note: Each row compares IGNNK with the strongest graph-based classical baseline for the corresponding dataset and masking ratio. Statistical significance is assessed using an exact paired permutation test with Benjamini–Hochberg false discovery rate correction across the nine node-missingness comparisons. Effect sizes are reported as the paired standardized mean difference (d_z), and confidence intervals denote bootstrap 95% confidence intervals for the mean MAE reduction. Mean absolute error (MAE), spatio-temporal graph signal processing (ST-GSP).

Table 6. Temporal-block completion on real sensor graphs: IGNNK against the strongest classical baseline, for all nine configurations

Dataset	Mask Frac.	Best Classical Baseline	Classical MAE	IGNNK MAE	Classical Advantage (%)
PEMS04	25%	temporal-linear	20.91 ± 0.16	27.08 ± 0.52	29.5
PEMS04	50%	temporal-linear	22.03 ± 0.07	28.81 ± 0.35	30.8
PEMS04	75%	temporal-linear	24.56 ± 0.21	31.37 ± 0.25	27.7
PEMS08	25%	temporal-linear	16.10 ± 0.13	21.70 ± 0.95	34.8
PEMS08	50%	temporal-linear	16.62 ± 0.17	22.15 ± 0.37	33.3
PEMS08	75%	temporal-linear	18.00 ± 0.17	24.01 ± 0.16	33.4
PeMSD7-M	25%	temporal-linear	1.95 ± 0.02	3.00 ± 0.08	53.9
PeMSD7-M	50%	temporal-linear	2.16 ± 0.03	3.12 ± 0.06	44.5
PeMSD7-M	75%	temporal-linear	2.56 ± 0.01	3.52 ± 0.04	37.3

Note: Temporal linear interpolation is the strongest classical baseline in all nine block-missingness configurations and outperforms IGNNK in every one of them, by 27.7% to 53.9% in MAE. The final column reports the percentage by which the classical baseline achieves lower MAE than IGNNK. Mean absolute error (MAE).

Overall, these experiments identify the specific regime in which graph learning provides operational value. The advantage is confined to whole-node, off-frame demand completion, precisely matching the deployment scenario addressed by the proposed framework. Outside this regime, classical interpolation methods remain the preferred solution and are therefore retained within the overall framework where appropriate.

6.5 R5-integrated architecture-aware case study on the constructed Prophet's Mosque graph

Experiment E5 evaluates the completion and analytic egress stages composed on the constructed Prophet's Mosque case-study graph. No learned parameters are carried over from the traffic benchmarks: the completion model is refitted from scratch on this graph, so what is reused is the methodology and not a trained model. Results are reported over six independent runs and evaluated against the analytic Oracle at four off-frame observability levels.

Necessity ladder. The principal result of the in-domain evaluation is the necessity ladder reported in Table 7, which compares three progressively richer pipeline configurations: CV-only, CV + analytic, and the CV + ST-GNN + analytic configuration.

The two transitions behave differently, and the interval formulation distinguishes their effects. The first transition (CV-only → CV + analytic) improves the guaranteed endpoint s_{min} by an order of magnitude at every observability level, because s_{min} depends on how the loaded network divides shared capacity and that information is available only once the architectural graph is introduced. Its effect on the ceiling endpoint s_{max} is smaller and operating-regime dependent: s_{max} is a per-zone quantity that does not involve contention, and at off-frame fractions of 0.6 and 0.8 the analytic configuration does not improve on the vision-only configuration for that endpoint. The vision-only configuration also cannot identify a binding bottleneck at all, so its binding-cut F1 is zero by construction rather than by measurement.

The second transition (CV + analytic → CV + ST-GNN + analytic) quantifies the contribution of off-frame demand completion, and it is the strongest result of the case study. Across the four observability levels the learned completion reduces s_{max} MAE from 0.0138-0.0933 to 0.0038-0.0046 and s_{min} MAE from 0.0264-0.1543 to 0.0116-0.0239, and raises binding-cut F1 from 0.80-0.98 to 0.99-1.00. The improvement holds at every off-frame fraction, at both endpoints, and at every horizon of the sensitivity grid except where a metric is saturated. Its magnitude increases as the observable portion of the venue decreases.

Component attribution. The analytic solver reproduces the Oracle solution exactly once the complete demand field is available, so the Oracle floor is zero by construction. Under the controlled perception-error proxy of relative magnitude 0.10 the perception term contributes only a small fraction of the total error; because that proxy is unbiased and more accurate than the counter measured in Experiment E1, this term is a lower bound rather than an estimate of deployed perception error. The dominant residual error arises from incomplete off-frame demand estimation, which the learned completion model reduces at every observability level and at both endpoints of the sufficiency interval.

Observability sweep. The influence of partial observation is examined by varying the off-frame fraction between 0.2 and 0.8.

Two complementary trends emerge. First, the benefit of the

learned completion is present at every observability level for both endpoints. Second, the operational gap between the analytic prior and the learned completion widens as observation degrades: for the ceiling endpoint it grows from 0.0092 to 0.0866 MAE, and for the guaranteed endpoint from 0.0148 to 0.1304 MAE, as the off-frame fraction rises from 0.2 to 0.8. The operational value of accurate off-frame demand estimation therefore increases substantially under severe partial observation.

Overall, the case study demonstrates that the same completion-and-egress methodology, refitted on the constructed Prophet's Mosque graph, operates as intended under simulated demand. The necessity ladder distinguishes the contributions of the analytic and completion stages, the component-attribution analysis identifies off-frame demand completion as the dominant source of residual error, and the observability study shows that the benefit of learned completion increases as a larger proportion of the venue becomes unobserved.

Finally, Figure 2 reports a robustness sweep across nine physically motivated demand regimes. The sweep uses a different occupancy generator from the experiment of Table 7, so the two are matched on the network loading regime ($\rho \approx 1.72$) rather than on the numerical value of T_{evac}, which corresponds to $T_{evac} \approx 198.8$ s for the sweep. The benefit of learned completion holds in seven of the nine regimes and the observability trend in eight. In the two remaining regimes the analytic configuration already attains zero s_{max} error, so no headroom is left for the completion stage; these are saturation cases rather than failures of the learned model.

Table 7. Necessity ladder on the constructed Prophet's Mosque graph (6 seeds, T_{evac} = 900 s)

Off-Frame Frac.	Pipeline Configuration	s_{max} MAE	S_{min} MAE	Binding-Cut F1
0.2	CV-only	0.0337 ± 0.0216	0.7825 ± 0.0208	0.000 ± 0.000
0.2	CV + analytic	0.0138 ± 0.0106	0.0264 ± 0.0207	0.979 ± 0.032
0.2	CV + ST-GNN + analytic	0.0046 ± 0.0003	0.0116 ± 0.0092	1.000 ± 0.000
0.4	CV-only	0.0584 ± 0.0220	0.7584 ± 0.0330	0.000 ± 0.000
0.4	CV + analytic	0.0440 ± 0.0132	0.0590 ± 0.0205	0.928 ± 0.032
0.4	CV + ST-GNN + analytic	0.0038 ± 0.0004	0.0178 ± 0.0108	0.993 ± 0.016
0.6	CV-only	0.0859 ± 0.0072	0.7448 ± 0.0205	0.000 ± 0.000
0.6	CV + analytic	0.0933 ± 0.0292	0.0622 ± 0.0148	0.921 ± 0.044
0.6	CV + ST-GNN + analytic	0.0043 ± 0.0007	0.0143 ± 0.0067	1.000 ± 0.000
0.8	CV-only	0.0868 ± 0.0292	0.7519 ± 0.0454	0.000 ± 0.000
0.8	CV + analytic	0.0911 ± 0.0365	0.1543 ± 0.0843	0.796 ± 0.134
0.8	CV + ST-GNN + analytic	0.0045 ± 0.0006	0.0239 ± 0.0162	0.993 ± 0.016

Note: Lower egress-sufficiency MAE and higher binding-cut F1 are better. s_{max} is the ceiling endpoint of Eq. (5); s_{min} is the guaranteed endpoint over all globally optimal allocations.Computer-vision (CV), spatio-temporal graph neural network (ST-GNN), mean absolute error (MAE).

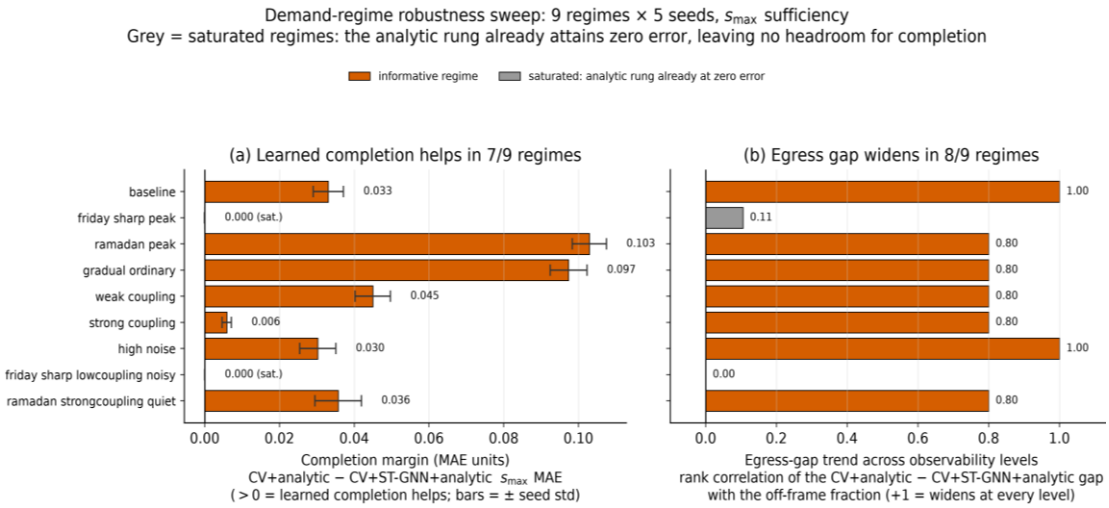


Figure 2. Demand-regime robustness sweep across nine regimes (5 seeds, 4 off-frame fractions), evaluated with the ceiling endpoint s_{max} at the reference loading $\rho \approx 1.72$

Note: Grey bars mark saturated regimes in which the analytic configuration already attains zero error. Panel (b) reports the rank correlation between the egress gap and the off-frame fraction across the four observability levels; it is a trend measure over observability and is unrelated to any ranking of venue zones.

6.6 Discussion

The experimental results support three principal conclusions. First, image observations alone are insufficient for operational egress reasoning, and explicit architectural topology is required in particular for the competition-sensitive component of egress sufficiency. Second, once the analytic network-flow formulation is introduced, off-frame demand completion becomes the dominant remaining source of residual error. Third, spatio-temporal graph learning provides measurable benefit specifically for whole-node (off-frame) demand completion, whereas classical interpolation remains preferable for conventional temporal gap filling. Collectively, these findings support the proposed decomposition of the problem into perception, demand completion, and analytic egress reasoning, within the evidential scope. The following subsections discuss the implications of these observations.

6.6.1 Interpreting the dual observability finding

Reconstruction accuracy and operational value behave

differently as observability decreases. The completion model holds a large reconstruction advantage at every off-frame fraction, while the operational benefit of the completed field grows as more of the venue becomes unobserved. These are not in tension: the reconstruction task is similar regardless of how many nodes are hidden, because each unobserved region is estimated from the same combination of spatial connectivity and temporal evolution.

The egress analysis, by contrast, depends on the quality of the alternative estimate available to the solver. As the observable portion shrinks, the non-learned prior degrades faster than the learned model, because a larger fraction of the field must be inferred from concurrent neighbouring observations alone without exploiting temporal correlation. The operational advantage therefore widens. Since the quantity of interest for crowd management is the quality of the resulting egress assessment rather than node-level accuracy alone, the value of completion is greatest under exactly the conditions the framework targets: substantial portions of the venue outside camera coverage.

6.6.2 What the necessity ladder establishes

The ladder is the clearest justification for including a learned component in an otherwise deterministic pipeline: the framework already contains a perception stage and an exact solver, so the learned component must supply something neither provides. The first transition is the empirical counterpart of the non-identifiability argument-bottleneck structure becomes expressible only once the analytic formulation is introduced, and the guaranteed endpoint becomes estimable only then-while the second isolates the value of completion. The results show that the first transition acts principally on the competition-sensitive endpoint, whereas the second improves both endpoints throughout.

The solver is exact once the demand field is available, but under partial observation its accuracy depends on the supplied field, which the component attribution analysis identifies as the dominant source of residual error. The ladder therefore supports the proposed division of labour: deterministic reasoning where the problem is exactly solvable, learning only for the latent off-frame demand field. Because the ladder is evaluated on the constructed case study, it establishes that division of labour for that setting rather than for deployed operation.

6.6.3 Operational interpretation and its limits

The framework produces zone-level egress-capacity sufficiency intervals and identifies the binding cut-edge set. Its objective is not to assign a generic crowd-risk score but to indicate which architectural elements limit pedestrian egress under the modeled assumptions. Rather than declaring a location unsafe, it identifies the cut edges that bound the clearable flow. Because the case-study graph is constructed from public information and has not been validated against official plans, on-site measurement, operational records or expert assessment, and because the demand field is simulated, the specific bottlenecks reported here are properties of the constructed model rather than verified bottlenecks of the venue. Any operational measure such an analysis might suggest-gate management, corridor reconfiguration, or flow control-is a candidate hypothesis requiring venue-specific validation before use.

Within the constructed case-study graph, the analytic solver localizes the binding cut to the gate-funnel structure at the venue boundary in the large majority of evaluated steps. This localization follows from the modeled topology and the assumed capacities, and is reported as a property of the case-study model rather than as a finding about the venue itself.

The necessity ladder demonstrates that reliable bottleneck identification depends on accurate off-frame demand completion. Although the analytic solver exactly identifies the critical cut once the demand field is available, inaccurate estimates of hidden demand can lead to incorrect bottleneck localization. The proposed spatio-temporal demand-completion model substantially improves agreement with the Oracle under partial observation, maintaining high binding-bottleneck accuracy even when a large proportion of the venue lies outside the observable camera coverage. Consequently, within this case study the learned component improves not only demand reconstruction but also the stability of the identified bottleneck structure under incomplete observation. The binding cut was found to be unique at every evaluated state of this constructed graph; that is an empirical property of the case-study model and is not claimed to hold for flow networks in general.

7. LIMITATIONS AND THREATS TO VALIDITY

The scope conditions and principal threats to validity are summarized below.

No target-venue imagery or measured demand. No registered Prophet's Mosque crowd imagery was quantitatively evaluated, and no measured zone-level pedestrian demand for the venue is publicly available. Zone-level demand in the case study is therefore simulated and the perception stage is represented by a controlled error proxy. The case study consequently demonstrates the proposed reasoning mechanism under controlled conditions; it does not measure deployment performance.

External benchmarks are methodological. The completion model is evaluated on real road-traffic sensor networks, which differ from pedestrian movement inside a religious venue in graph structure, signal semantics and temporal dynamics. Those experiments establish that the completion methodology works on real graph-structured spatio-temporal measurements; they do not establish pedestrian-demand reconstruction accuracy at the Prophet's Mosque. No trained weights are transferred: the same methodology is re-instantiated and refitted on the target graph.

Image-to-graph registration. Mapping image observations to graph nodes requires accurate camera registration. The registration component is provided as an illustrative implementation and was not evaluated on venue imagery; registration error and oblique viewing geometry may reduce the accuracy of both the perception stage and the downstream demand estimates. Quantitative evaluation on registered mosque imagery remains future work.

Unvalidated case-study graph and capacities. The venue graph was constructed from publicly available information and was not validated against official architectural or engineering plans, on-site measurements, operational crowd-management records, or expert assessment. Most edge widths are estimated rather than surveyed, and edge capacities follow from those widths and a fixed specific-flow assumption. Because every capacity scales linearly with the specific-flow parameter, a global change rescales all cuts identically and leaves the identity of the binding cut unchanged, while the sufficiency

endpoints scale with it. Absolute capacity values, by contrast, depend directly on these assumptions. Reported bottlenecks and capacities are therefore conditional on the constructed model.

Perception bias. The CV front-end exhibits the density-dependent undercount bias identified in Experiment E1. Consequently, at very high crowd densities, underestimated demand may lead to overestimated egress-capacity sufficiency. This reinforces the need for target-domain calibration before operational use, although it does not affect the qualitative conclusions regarding off-frame demand completion or analytic egress reasoning.

Static formulation. The egress model is a static network-capacity formulation. It does not represent pedestrian interactions, counterflows, queue formation or spillback, congestion propagation, crowd turbulence, behavioral response, dynamic route choice, or time-dependent evacuation. Egress-capacity sufficiency therefore bounds what the network geometry can clear; it does not by itself establish operational crowd safety or detect hazardous conditions.

No uncertainty quantification. All reported quantities are point estimates. The framework does not produce a calibrated perception uncertainty, and no coverage analysis or predictive interval is reported for egress-capacity sufficiency. The perception-error perturbations used in Experiments E2 and E5 are controlled sensitivity settings rather than uncertainty quantification. Calibrated uncertainty estimation and its propagation through the completion and egress stages remain future work.

Constructed comparisons. Several reported comparisons are controlled constructions rather than independent empirical evidence: the topology-hard matched-pair regime fixes visible pedestrian demand by design, the analytic baseline reproduces the Oracle by construction, and the vision-only configuration cannot express a cut, so its binding-cut F1 is zero by construction. These are reported as mechanism demonstrations and exactness checks. The independent empirical evidence in this paper is the perception evaluation on real imagery and the completion evaluation on real traffic networks, including its negative outcomes under block and spatial-only missingness.

8. REPRODUCIBILITY

The implementation covers graph construction, the analytic solver, the CV evaluation, the completion models, the classical baselines and the evaluation scripts.

The Prophet's Mosque case-study graph and its capacity configuration, including per-edge width provenance flags, are included as reproducible artifacts. Crowd imagery is referenced through the original data sources rather than redistributed, while all derived annotations required for the experiments are provided. The image-to-node registration component is released as an illustrative implementation and is not accompanied by venue registration data.

The analytic egress labels are deterministic and independently verifiable: given the released graph and edge capacities, the reported binding cut and egress-sufficiency intervals can be reproduced directly from the network-flow formulation. The global maximum-flow value and the minimum-cut value are well-defined, and the binding cut-edge set was unique at every evaluated state of this graph. The division of that flow among zones, however, need not be unique, so no single per-zone allocation is reported; the endpoints $s_{min}(z)$ and $s_{max}(z)$ are computed instead and are independent of the maximum-flow solution a solver returns.

The source code, experimental artifacts, and instructions required to reproduce the experiments, tables, and figures reported in this paper will be made publicly available upon acceptance. The demand-regime robustness sweep of Figure 2 uses a separate demand generator from the experiment of Table 6; both were evaluated under the same sufficiency definition at a matched network loading, and the per-regime outputs are released alongside the code.

The datasets analyzed in this study are publicly available third-party benchmarks (ShanghaiTech Part B, PEMS04, PEMS08, and PeMSD7-M) that are referenced through their original sources and are not redistributed, and the reproducibility artifacts developed in this study are available from the corresponding author upon reasonable request.

9. CONCLUSIONS

This paper presented an architecture-aware framework for reasoning about egress-capacity sufficiency under partial observation. The framework integrates CV visible-demand estimation, spatio-temporal graph neural networks for off-frame demand completion, and deterministic max-flow/min-cut analysis of static network-level egress capacity.

The evaluation produced three findings. First, visible image observations alone are insufficient for egress reasoning, and architectural topology is required in particular for the competition-sensitive component of egress sufficiency. Second, spatio-temporal graph learning provides a measurable advantage specifically for whole-node (off-frame) completion, whereas classical interpolation remains preferable for temporal gap filling and no advantage is observed in the spatial-only setting. Third, combining off-frame completion with analytic network-flow reasoning improves bottleneck identification and egress-capacity assessment under partial observation within the constructed case study.

The framework was evaluated with real CV models on public crowd imagery, real traffic benchmarks for off-frame completion, and a constructed architectural case study of the Prophet's Mosque with simulated zone-level demand. No registered venue imagery or measured occupancy was evaluated, and the case-study graph is not an officially validated model of the venue; the results therefore support the proposed decomposition into perception, demand completion, and analytic egress reasoning, while bounding the scope of the resulting claims.

Future work will focus on validating the venue graph against authoritative architectural records, evaluating the framework with measured occupancy data, introducing calibrated uncertainty estimation and its propagation to egress-capacity sufficiency, improving image-to-graph registration on real mosque imagery, extending the analytic formulation to dynamic evacuation models, and applying the same partial-observation reasoning template to other capacitated-venue domains. More broadly, this work demonstrates that architecture-aware operational reasoning provides a principled foundation for crowd-safety assessment under partial observation, extending computer vision beyond density estimation toward infrastructure-aware operational decision support.

REFERENCES

[1] Feliciani, C., Corbetta, A., Haghani, M., Nishinari, K. (2023). Trends in crowd accidents based on an analysis of press reports. Safety Science, 164: 106174. https://doi.org/10.1016/j.ssci.2023.106174

[2] Helbing, D., Johansson, A., Al-Abideen, H.Z. (2007). Dynamics of crowd disasters: An empirical study. Physical Review E, 75(4): 046109. https://doi.org/10.1103/PhysRevE.75.046109

[3] Liang, H.Y., Lee, S., Sun, J., Wong, S.C. (2024). Unraveling the causes of the Seoul Halloween crowd-crush disaster. PLOS ONE, 19(7): e0306764. https://doi.org/10.1371/journal.pone.0306764

[4] Jia, X.L., Feliciani, C., Murakami, H., Nagahama, A., Yanagisawa, D., Nishinari, K. (2022). Revisiting the level-of-service framework for pedestrian comfortability: Velocity depicts more accurate perceived congestion than local density. Transportation Research Part F: Traffic Psychology and Behaviour, 87: 403-425. https://doi.org/10.1016/j.trf.2022.04.007

[5] Berggold, P., Borrmann, A., Nousias, S. (2025). Graph neural networks for building evacuation prediction: Integrating real-time pedestrian flow simulation into BIM workflows. In EG-ICE 2025: Proceedings of the 32nd EG-ICE International Workshop on Intelligent Computing in Engineering. University of Strathclyde Publishing, Glasgow, pp. 231-239. https://doi.org/10.17868/strath.00093235

[6] Zhang, Y.Y., Zhou, D.S., Chen, S.Q., Gao, S.H., Ma, Y. (2016). Single-image crowd counting via multi-column convolutional neural network. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 589-597. https://doi.org/10.1109/CVPR.2016.70

[7] Li, Y.H., Zhang, X.F., Chen, D.M. (2018). CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 1091-1100. https://doi.org/10.1109/CVPR.2018.00120

[8] Ma, Z.H., Wei, X., Hong, X.P., Gong, Y.H. (2019). Bayesian loss for crowd count estimation with point supervision. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 6141-6150. https://doi.org/10.1109/ICCV.2019.00624

[9] Wang, B.Y., Liu, H.D. Samaras, D., Hoai, M. (2020). Distribution matching for crowd counting. arXiv Preprint arXiv: 2009.13077. https://doi.org/10.48550/arXiv.2009.13077

[10] Idrees, H., Tayyab, M., Athrey, K., et al. (2018). Composition loss for counting, density map estimation and localization in dense crowds. In European Conference on Computer Vision (ECCV). Springer, Cham, pp. 544-559. https://doi.org/10.1007/978-3-030-01216-8_33

[11] Sindagi, V.A., Yasarla, R., Patel, V.M. (2022). JHU-CROWD++: Large-scale crowd counting dataset and a benchmark method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(5): 2594-2609. https://doi.org/10.1109/TPAMI.2020.3035969

[12] Wang, Q., Gao, J.Y., Lin, W., Li, X.L. (2021). NWPU-Crowd: A large-scale benchmark for crowd counting and localization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(6): 2141-2149. https://doi.org/10.1109/TPAMI.2020.3013269

[13] Oh, M.H., Olsen, P., Ramamurthy, K.N. (2020). Crowd counting with decomposed uncertainty. AAAI Conference on Artificial Intelligence, 34(7): 11799-11806. https://doi.org/10.1609/aaai.v34i07.6852

[14] Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine, 30(3): 83-98. https://doi.org/10.1109/MSP.2012.2235192

[15] Ortega, A., Frossard, P., Kovačević, J., Moura, J.M.F., Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. Proceedings of the IEEE, 106(5): 808-828. https://doi.org/10.1109/JPROC.2018.2820126

[16] Yu, H.F., Rao, N., Dhillon, I.S. (2016). Temporal regularized matrix factorization for high-dimensional time series prediction. In Advances in Neural Information Processing Systems 29 (NIPS 2016). Curran Associates, Red Hook, pp. 847-855.

[17] Wu, Y.K., Zhuang, D.Y., Labbe, A., Sun, L.J. (2021). Inductive graph neural networks for spatiotemporal kriging. Proceedings of the AAAI Conference on Artificial Intelligence, 35(5): 4478-4485. https://doi.org/10.1609/aaai.v35i5.16575

[18] Cini, A., Marisca, I., Alippi, C. (2022). Filling the G_ap_s: Multivariate time series imputation by graph neural networks. arXiv Preprint arXiv: 2108.00298. https://doi.org/10.48550/arXiv.2108.00298

[19] Ford, L.R., Fulkerson, D.R. (1962). Flows in Networks. Princeton University Press, Princeton. https://doi.org/10.1515/9781400875184

[20] Ahuja, R.K., Magnanti, T.L., Orlin, J.B. (1993). Network Flows: Theory, Algorithms, and Applications. Prentice Hall, Englewood Cliffs.

[21] Skutella, M. (2009). An introduction to network flows over time. In Research Trends in Combinatorial Optimization, Springer, Berlin, pp. 451-482. https://doi.org/10.1007/978-3-540-76796-1_21

[22] Fruin, J.J. (1971). Pedestrian Planning and Design. Metropolitan Association of Urban Designers and Environmental Planners, New York.

[23] Abouelaziz, I., Ghalmane, Z. (2026). DRAGON: A dynamic risk-aware graph optimization network for adaptive building evacuation using graph convolutional network and Q-learning. Multimedia Tools and Applications, 85: 197. https://doi.org/10.1007/s11042-026-21400-9

[24] Chang, X.J., Ren, P.Z., Xu, P.F., Li, Z.H., Chen, X.J., Hauptmann, A. (2023). A comprehensive survey of scene graphs: Generation and application. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1): 1-26. https://doi.org/10.1109/TPAMI.2021.3137605

[25] Lee, J., Li, K.J., Zlatanova, S., et al. (2020). OGC® IndoorGML 1.1. Open Geospatial Consortium, Arlington. https://doi.org/10.62973/19-011r4