

A Neural Field-Based Approach for View Computation & Data Exploration in 3D Urban Environments

Stefan Cobeli , Kazi Shahrkhan Omar , Rodrigo Valen  a, Nivan Ferreira , and Fabio Miranda

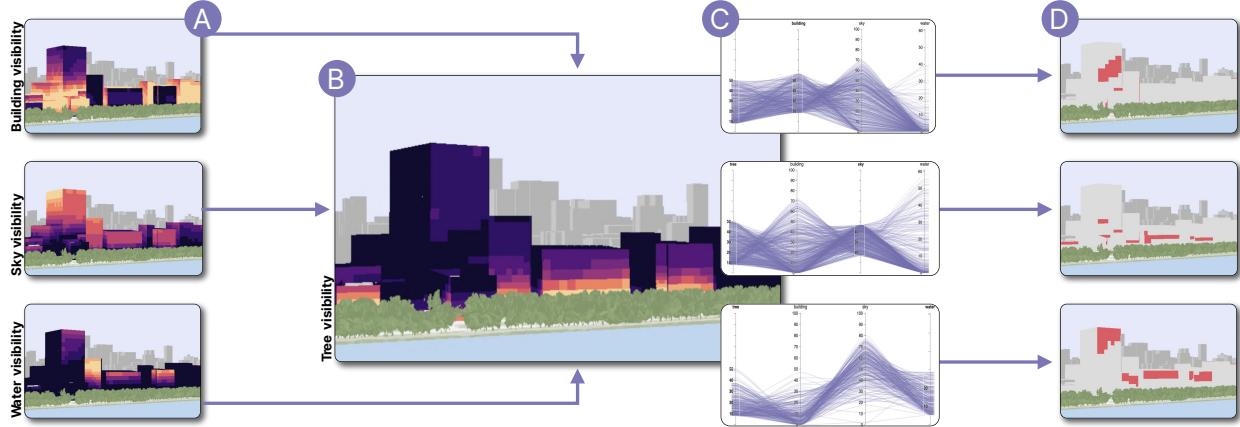


Fig. 1: Direct queries for the computation of thematic data associated with building facades representing **A** the visibility towards buildings, sky, and water, and **B** the visibility towards trees. Following the user specified conditions in **C** the parallel coordinates we perform **D** inverse queries to infer the building facade positions satisfying the joint visibility constraints.

Abstract—Despite the growing availability of 3D urban datasets, extracting insights remains challenging due to computational bottlenecks and the complexity of interacting with data. In fact, the intricate geometry of 3D urban environments results in high degrees of occlusion and requires extensive manual viewpoint adjustments that make large-scale exploration inefficient. To address this, we propose a view-based approach for 3D data exploration, where a vector field encodes views from the environment. To support this approach, we introduce a neural field-based method that constructs an efficient implicit representation of 3D environments. This representation enables both faster *direct queries*, which consist of the computation of view assessment indices, and *inverse queries*, which help avoid occlusion and facilitate the search for views that match desired data patterns. Our approach supports key urban analysis tasks such as visibility assessments, solar exposure evaluation, and assessing the visual impact of new developments. We validate our method through quantitative experiments, case studies informed by real-world urban challenges, and feedback from domain experts. Results show its effectiveness in finding desirable viewpoints, analyzing building facade visibility, and evaluating views from outdoor spaces. Code and data are publicly available at urbantk.org/neural-3d.

Index Terms—Urban computing, urban visual analytics, view computation, neural field

1 INTRODUCTION

The increasing availability of urban datasets with 3D spatial components has created new opportunities for analyzing complex urban phenomena [28]. Among these possibilities are applications such as visibility evaluation [50], view analysis of environmental features [46, 52, 54], visibility planning [12, 36], view preferences [18, 38], and building design [10, 43], which often rely on computing view indices or identifying viewpoints that satisfy specific criteria. Conventional methods for computing these indices typically rely on ray casting [27] or rasterization [12], both of which are computationally expensive and scale poorly to large urban settings. To mitigate this, many studies either perform offline analyses [12, 28, 40] or rely on pre-computation to enable interactive exploration [10]; however, both approaches ultimately limit true interactivity in visual analysis. Beyond computation, exploring and interacting with 3D urban datasets presents additional difficulties. Spatial occlusion and complex camera controls often hinder interpretation:

once indices are obtained, they are commonly visualized as thematic overlays on physical surfaces (Fig. 1(A,B)). Unlike 2D visualizations with fixed perspectives, 3D exploration requires careful viewpoint selection and manipulation across multiple degrees of freedom. Without appropriate support, users may struggle to efficiently navigate, reveal patterns, and interpret the data. Although prior research has extensively explored methods for navigating and visualizing 3D urban environments, much of this work has focused on the representations of cities rather than the interactive analysis of 3D urban data itself [19]. For instance, recent studies have investigated viewpoint selection [3], occlusion-aware rendering [1], and deformation-based de-occlusion and navigation support [5]. However, these approaches only consider physical data, such as building geometry and infrastructure, while overlooking the integration of thematic attributes. As a result, two intertwined challenges arise: (1) the **efficient computation** of view-related measures that capture the relationships between physical elements (e.g., building geometry, terrain) and thematic attributes (e.g., sunlight access, visibility), and (2) the **effective exploration** of such measures in 3D environments, where interaction and interpretation are inherently more complex than in 2D settings.

To address these challenges, we propose a view-based approach for the interactive exploration of physical and thematic dimensions in this work. At the core of our method is the modeling of views as a dense multi-dimensional field, where each viewpoint is associated with a semantic attribute vector that encodes relevant properties across large

• Stefan Cobeli, Kazi Shahrkhan Omar, and Fabio Miranda are with the University of Illinois Chicago. E-mail: {scobel2,komar3,fabiom}@uic.edu
• Rodrigo Valen  a and Nivan Ferreira are with the Universidade Federal de Pernambuco. E-mail: {rbv3,nivan}@cin.ufpe.br.

urban environments and provides a unified representation for analyzing spatial patterns. Our proposal specifically targets mesh-based urban representations, where surfaces can be directly associated with thematic attributes. This representation enables users to assess variations in data distribution, identify patterns, and make comparisons within the urban environment. To efficiently support the operations described above, we introduce a neural field-based approach, inspired by the well-established neural radiance fields [26], where a deep neural network learns the vector field. In fact, this model enables estimating the vector field value (i.e., semantic view assessment indices) without using additional rendering or ray casting steps. Furthermore, the differentiable nature of this neural representation allows for the introduction of an algorithm for identifying viewpoints that follow criteria defined by the thematic data, which eliminates the need for explicit field computation across each viewpoint, leveraging the network’s ability to approximate the field efficiently. In addition to enabling efficient exploration of 3D urban data, our method can be leveraged for a wide range of applications. By encoding physical and thematic attributes in the vector field, the approach supports the analysis of sightlines, occlusion, and visibility across large urban areas.

To evaluate our approach, we conducted a set of quantitative experiments to assess its performance and accuracy. Specifically, we measured how quickly views are generated and how accurately they capture the intended spatial relationships within the urban environment. Beyond these assessments, we have implemented a visual interface (see Fig. 1) and through it we present three case studies to demonstrate the applicability of our approach in addressing key challenges faced by urban experts. First, we illustrate how our method can be used to discover novel viewpoints by searching for custom visibility distributions. Second, we present an efficient building facade visibility analysis with views from buildings towards the outdoors, supporting urban planning and architectural design considerations. And third, we employ our method to explore views from outdoor spaces towards buildings, an important aspect for urban aesthetics and environmental assessments. Finally, we report on the feedback on our model received in a series of interviews with domain experts. Code and data are publicly available at urbantk.org/neural-3d. Our contributions can be summarized as follows:

- We present a novel approach that models *views* as a vector field encoding both physical and thematic urban attributes.
- We propose a neural field-based method that efficiently learns and approximates the vector field.
- We demonstrate the versatility of our model through experimental evaluation, three practical usage scenarios, and expert feedback received in a series of interviews.

2 RELATED WORKS

This section reviews previous research related to various aspects of this work. In particular, we review contributions and challenges in visual analytics in 3D urban environments, navigation and occlusion management techniques, and viewpoint estimation.

2.1 Visual analytics for 3D urban data

Visual analytics for 3D urban data is a research field motivated by the fact that many phenomena of interest in urban studies are inherently three-dimensional, requiring reasoning over the 3D structure of urban environments. For this reason, applications in domains like civil engineering, urban planning, architecture, and climate science involve not only data processing of 3D urban data but also using 3D visual representations to properly represent the spatial context of the studied data. In these scenarios, thematic data layers (e.g., sunlight access, visibility scores) are associated with physical ones (e.g., building geometry, streets, terrain). As extensively discussed in recent works by Mota et al. [33] and Miranda et al. [28], unlike the 2D case, there is a lack of visual metaphors designed explicitly for 3D urban environments, as well as a scarcity of evaluation studies that offer grounded design and usage guidelines for such metaphors. In fact, incorporating the third dimension significantly increases the complexity of designing effective interactive visual representations. First, features such as perspective

distortion and aesthetic visual effects like shading can interfere with critical visual channels commonly used for data representation. Second, and more relevant to this work, the typical spatial arrangement and density of physical structures in urban environments lead to significant occlusion. Altogether, these factors contribute to the complexity of data exploration in 3D urban settings.

Elmqvist and Tsigas’s taxonomy [11] describes the different modes of occlusion and categorizes de-occlusion methods. Building on their work, Miranda et al. [28] categorized occlusion handling methods applied to 3D urban analytics into six categories: *bird’s view*, *multi-view*, *ghosting*, *deformation*, *slicing*, and *assisted steering*. Despite the diversity of methods this classification suggests, these rarely find their way into 3D urban analytics solutions. In fact, most previous works rely on the users taking a bird’s view through manual camera manipulation to navigate the environment. The other categories consist of providing a multi-perspective view (*multi-view*), removing occluding parts via transparency (*ghosting*), deforming physical data layers (*deformation*), and slicing volumetric datasets (*slicing*). Ghosting techniques are limited due to the difficulty in handling multiple layers of occluders [9]. Deng et al. [9] proposed a hybrid approach that applies various deocclusion operators to a given viewpoint. The operators include bird’s-eye view adjustments (increasing the camera height) and deformation techniques such as road shifting, building scaling, and building displacement.

On the other end of the spectrum, assisted steering techniques were found to be the least prevalent in 3D urban visualization systems. These techniques automatically compute optimal viewpoints according to some criteria. The solution proposed in our paper falls in this category. As Miranda et al. [28] surveyed, the assisted steering methods applied to 3D urban environments have two main limitations. First, they only consider physical layers in their optimality criteria. Second, they can be computationally expensive, particularly when relying on non-differentiable viewpoint optimality criteria [3]. Our proposed method consists of a deep learning-based approach that enables fast and flexible selection of viewpoints based on thematic data layers.

2.2 Visibility exploration in 3D urban environments

Many essential use cases are targeted by 3D urban visual analytics studies [28]. Among them, visibility analysis is very relevant to our work. Domain works focusing on this area tackle problems like view access equity [50], visibility to green spaces [46, 52], blue spaces [54], and landmarks [40], views of high rises [21], and office rents [44]. In addition, some works in the visualization community focused on designing visual analytics systems for visibility analysis. For example, the works by Ortner et al. [36] and Ferreira et al. [12] presented design studies to support visibility-aware urban planning and view impact analyses, respectively. These works relied on a rasterization strategy to compute visibility scores. However, since visibility computation is expensive, these could impact the system’s interactivity when operations requiring many scene renderings are executed. For this reason, different strategies were proposed. For example, Doraiswamy et al. [10] proposed a computational topology-based framework for building design. This work relied on a pre-computation-based strategy to support interactive visualization and real-time building design optimization. The results of the pre-computation were stored in a regular grid data structure. Closer to our work, Li et al. [21] used a deep learning-based model to estimate visibility-based scores to assess the quality of the views from the windows in building facades. While close in intent, they employ a different modeling strategy than ours. In fact, they perform a pixel-level segmentation step followed by a regression phase to compute the visibility scores. Our model reduces this computation by directly computing the semantic aggregation from the views (without the segmentation step). Finally, unlike previous works, our model is used not only to accelerate visibility score computation but also to enable semantic search over viewpoints.

2.3 Viewpoint selection techniques

Viewpoint selection has been a widely studied problem in different fields. When view targets are clearly defined, direct methods can be

used to place the camera to maximize views towards this target. These methods consist of heuristics that compute camera parameters from the target's spatial position and extent. For example, Li et al. [20] introduced a 3D visualization technique for trajectory attribute data, combining it with a geometric viewpoint selection strategy to either focus on specific trajectory segments or provide an overview with a broader viewing area. In more complex scenarios, *optimization-based methods* are the most common. This approach involves defining a quality measure for a given viewpoint and using optimization algorithms to select the best viewpoint based on that measure. Bonaventura et al. [2] presented a survey of viewpoint selection quality measures for polygonal models, primarily targeting computer graphics applications. The measures analyzed focus on assessing geometric properties (e.g., area, silhouette, depth, surface curvature) of a polygonal mesh as perceived from a given viewpoint.

Few works have addressed viewpoint selection in 3D urban environments. Li et al. [22] followed the direct approach to generate customized and narrative animations in an evacuation planning context. Neuville et al. [35] proposed a technique that uses visibility to user-defined points of interest as a quality measure. Viewpoints are sampled in a hemisphere around the points of interest, and the quality measure is computed for all these points to assess the best one. Zhang et al. [53] and Boorboor et al. [3] built an immersive analytics system for site development and flood simulation analysis, respectively. They followed the optimization-based approach to select viewpoints towards a set of user-defined points of interest. This was done by optimizing an energy function that considers geometrical visibility and spatial location aspects. In all of these techniques, the viewpoint selection criteria only considered physical elements of the city and also focused on specific geometric elements (e.g., buildings) as targets. Our technique differs from these in two key ways. First, it is designed with a different goal in mind. More precisely, we are interested in assisting the exploration of *semantic view indices*, i.e., views considering elements of thematic data layers. In our approach, the user specifies semantic elements of what they desire to visualize instead of physical targets. Second, we avoid expensive visibility computations by using a neural network approximating the visibility given a viewpoint. As detailed in Section 4, our model is trained on a relatively small set of viewpoints and approximates the desired visibility function well.

Other works also approached the view selection problem via machine learning methods. These were proposed mainly in the computer vision community [16]. Sun et al. [41] proposed optimizing multi-camera placement for maximum visibility coverage. Their method approximates visibility coverage using a neural network and reformulates the multi-camera placement problem into a continuous optimization task. While their contributions are similar to ours, we adopt a different modeling strategy inspired by neural radiance fields. To the best of our knowledge, our work represents the first machine learning-based approach applied specifically to 3D urban visual analytics. Moreover, our highly flexible model supports evaluating a wide range of previously proposed metrics and enabling diverse analysis tasks.

3 BACKGROUND & CHALLENGES

In this paper, we investigate a new approach for two intertwined problems in 3D urban data exploration: view estimation and visibility analysis. 3D urban data is often derived from sources such as LiDAR scans, simulations, and modeling, and it provides a representation of the urban environment, capturing both physical and thematic attributes relevant to buildings and infrastructure. These datasets are widely used in urban planning [39], environmental assessments [14], and infrastructure planning [29], but their complexity makes extracting meaningful insights challenging. Our recent survey summarized a series of domains, tasks, and data types associated with 3D urban analysis, highlighting key challenges in effectively leveraging these datasets [28]. One fundamental aspect is the interplay between physical and thematic data. Physical data, such as the geometry of buildings and infrastructure, define the structural environment, while thematic data, often derived from simulations, encodes additional attributes like solar exposure, air quality, or noise levels. This interplay leads to occlusion problems, as thematic

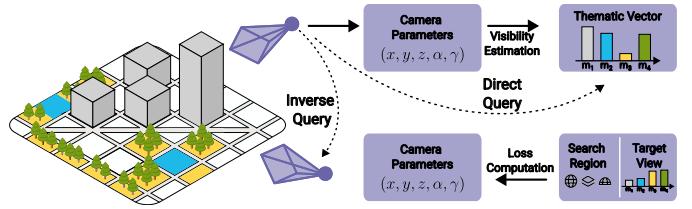


Fig. 2: Our proposed visibility-based exploration is enable by two types of queries supported by a machine learning model. Direct queries (top), computing a thematic vector from spatial positions. Inverse queries (bottom), searching spatial positions satisfying given target view constraints.

overlays are often projected onto physical layers (e.g., building surfaces), making them difficult to analyze when obstructed from certain viewpoints. However, both physical and thematic data are essential for analyses [28], meaning that effective exploration must account for their spatial relationships and visibility constraints.

Visualization and visual analytics systems that leverage such data often rely on manual camera navigation to resolve these issues [10, 12, 27, 47], which means that the user must adjust viewpoints iteratively to reveal occluded thematic information. This is cumbersome and time-consuming. Unlike 2D visualizations, 3D environments demand constant interaction (panning, zooming, rotating) to explore structures of the physical and thematic overlays, with users often resorting to trial-and-error navigation to access relevant data. In a large-scale urban environment, such an approach is particularly inefficient, as relevant data may be scattered across multiple locations. View estimation can address this problem by determining optimal viewpoints, supporting a more effective exploratory process while reducing reliance on exhaustive manual adjustments. Fast, automated view estimation can enhance interactive analysis by guiding users toward informative perspectives.

Parallel to this, visibility analysis is a domain-specific problem that focuses on analyzing the visibility of urban structures and natural features from a set of viewpoints. Thus, visibility analysis requires the computation of new data that quantifies which elements of the built environment are visible from given locations. This process has applications in solar exposure assessment [13], urban design evaluations [48], and sightline analysis [6]. However, traditional visibility analysis techniques, such as exhaustive ray casting, are computationally expensive and scale poorly to large urban settings. This process can also benefit from an effective view estimation computation.

Our approach integrates fast view estimation as a means to enhance both data exploration and visibility analysis. At its core, we create an implicit representation of the 3D environment. Such representation contains, for a given position, an estimate of the data distribution seen from that position, incorporating both physical and thematic attributes. By employing an implicit approach, we are able to perform fast computation of requested viewpoints to support interactive data exploration.

4 VIEW-BASED 3D URBAN DATA EXPLORATION

Motivated by the aforementioned challenges and the growing availability of 3D urban data, we propose an alternative method for exploring these datasets, shifting away from traditional navigation-based exploration towards a view-driven, queryable framework. To support this view-driven framework, at the core of our work is the need to *estimate views*, i.e., compute a distribution capturing both physical and thematic data visible from a given viewpoint. Such estimation of views will support an exploration of the 3D urban environment without the need for manual navigation. To achieve this, our method builds a neural implicit representation of the 3D environment, where per-view data distributions are encoded as a continuous vector-valued function. This representation supports: (1) **direct queries**, where users retrieve metrics for any viewpoint without requiring scene navigation or expensive ray tracing, and (2) **inverse queries**, where users specify a desired data pattern and retrieve viewpoints that satisfy these criteria. Our approach

adopts a methodology inspired by the well-established neural radiance fields [26]. However, instead of predicting radiance and opacity along a camera ray to synthesize an image per viewpoint, our model learns to approximate the distribution of data metrics per viewpoint, capturing both physical visibility and thematic attributes of the environment. In other words, instead of reconstructing an image, we learn a continuous function that maps spatial coordinates and viewing directions to data distributions. This allows us to retrieve information about a scene rather than reconstruct its appearance (see Fig. 2). Next, we detail our approach, first covering our vector field representation (Section 4.1) followed by our model to learn an implicit representation of this vector field (Section 4.2).

4.1 Vector-valued function representation

We define a viewpoint as a tuple $(x, y, z, \alpha, \gamma) \in \mathbb{R}^5$, where (x, y, z) denotes the 3D spatial position in Cartesian coordinates and (α, γ) is a viewing direction represented by Euler angles (we ignore roll). The set of all viewpoints is denoted by \mathcal{V} . We assume that the physical layers describing the urban environment are represented as polygonal meshes. These include both nature elements (e.g., trees, bodies of water) and man-made elements (e.g., roads, buildings). Furthermore, we assume that we are given a thematic data layer represented by a vector field T defined on the vertices of the mesh representing the elements of our physical layer. This vector field sets the semantic context for our analysis and depends on the specific application. For example, in a sunlight access study, T might represent the amount of direct sunlight received on a building surface. In a visibility study, we could classify the different physical elements into categories (e.g., buildings, roads, trees, and water) and assign via T an integer value for each category.

To represent the 3D urban environment, we define a vector-valued function representation that encodes both spatial structure and view-dependent thematic data. We call this function $S : \mathcal{V} \rightarrow \mathbb{R}^k$, where $S(x, y, z, \alpha, \gamma) \in \mathbb{R}^k$ produces a vector $\mathbf{m} = (m_1, m_2, \dots, m_k)$ that encodes k scalar values describing what can be observed from that viewpoint. To define the value of \mathbf{m} for a given viewpoint, we first rasterize the 3D scene, producing a 2D image where each pixel encodes the physical elements and the thematic vector of view indices T from the surfaces visible from that viewpoint. The image will provide a visual summary of the environment as perceived from the given viewpoint. After rendering the rasterized image, we extract a distribution of view indices by aggregating pixels into bins. In fact, each component m_i of \mathbf{m} represents the aggregated count of pixels falling into a specific range within the view indices. Formally, we denote by $I := I_{(x,y,z,\alpha,\gamma)}$ the rasterized image for a given viewpoint, and (u, v) represents the pixel coordinates, $I(u, v) \in \mathbb{R}$ is a thematic attribute observed in that pixel. The distribution of view indices will be given by $m_i = \sum_{(u,v)} \mathbb{1}_{B_i}(I(u, v))$, where m_i is the aggregated value for a bin B_i , considering a set of bins \mathbf{B} , and $\mathbb{1}_{B_i}$ is the indicator function of the bin B_i . The sum accumulates the thematic data within each bin across all pixels. Finally, the counts in the bins are normalized so the sum of the values adds up to 1.

For example, in a sunlight access study, the thematic value T might represent the amount of direct sunlight received on a building surface. In this scenario, we could define three bins (B_1 for 0-2 hours of sunlight, B_2 for 2-5 hours, and B_3 for five or more hours) and the resulting feature vector $\mathbf{m} = (m_1, m_2, m_3)$ will indicate the proportion of visible surfaces that belong to each bin. Such representation naturally encodes contributions to a given thematic attribute, taking into account visibility constraints. Moreover, by taking into account the distribution of thematic values across viewpoints, this representation allows us to retrieve camera positions that meet certain filtering constraints.

Since S encodes distributions for a large number of positions and viewing directions in a 3D environment, it results in a high-dimensional and large-scale representation that might span large areas with millions of potential viewpoints. Therefore, storing and processing this function explicitly becomes impractical. To address this computational challenge, we propose a compact and efficient way to store and query S without explicitly maintaining the original data. Next, we detail how we use an implicit neural representation, which learns a function

approximation of S , reducing storage usage while enabling efficient retrieval of thematic distributions.

4.2 Neural field-based implicit representation

To represent the vector-valued function S without explicitly storing all values, we adopt an implicit neural representation inspired by neural radiance fields. This representation provides a compressive representation of the visibility function S with a much smaller memory footprint than we would need to store the raw data [24]. However, unlike traditional neural radiance fields, which model radiance and opacity to synthesize images, we approximate the distribution of thematic values visible from any given viewpoint in the 3D environment. In other words, instead of storing discrete values, we train a neural network $F_\Theta : \mathcal{V} \rightarrow \mathbb{R}^k$ that will approximate the original vector field.

4.2.1 Model architecture

F_Θ is implemented as a multi-layer perceptron that takes as an input the spatial position and viewing direction and outputs the predicted distribution of view indices, i.e., $\mathbf{m} = F_\Theta(x, y, z, \alpha, \gamma)$.

To start, each coordinate and viewing direction is embedded using a high-frequency function, which was previously used in the neural radiance fields methodology. This type of input embedding was shown to improve the performance of neural networks when learning high-variation data. Specifically, we project the input using $\rho : \mathbb{R} \rightarrow \mathbb{R}^{10}$, $\rho(t) = (\sin(2^0 \pi t), \cos(2^0 \pi t), \dots, \sin(2^9 \pi t), \cos(2^9 \pi t))$. The input obtained for the 3D coordinates $(\rho(x), \rho(y), \rho(z))$ is passed through 10 fully-connected layers, each with 256 nodes and with ReLU activations. We included a skip connection by concatenating the first layer’s inputs with the sixth layer’s input. We used a 128-node layer with a hyperbolic tangent activation function as a final layer. This takes as input the direction embeddings $\rho(\gamma), \rho(\theta)$ and the output of the 10-layer network to return the thematic distribution \mathbf{m} .

4.2.2 Generalized model

We generalize the model described above by introducing two additional modules: input regularization and output customization. First, we show how the input can be restricted to a regular subspace. As our input domain is \mathbb{R}^3 and our scene is bounded, we can consider that it actually resembles an irregular 3D cube. Learning subspaces is an extension of learning the values of the function S in the whole domain. To exemplify, we consider bounded 2D hyperplanes as parallelograms contained in the cube domain. We define a parallelogram by its origin p , its normalized side direction vectors (v_1, v_2) , and lengths (l, L) . These constants induce a parametrization for every point (x, y, z) in the given parallelogram, by finding an $a, b \in [0, 1]$ such that $(x, y, z) = \zeta(a, b)$:

$$\zeta(a, b) = p + alv_1 + blv_2.$$

Following the same steps, the parametrization function ζ can be replaced by any other parameterizable spatial region in the 3D cuboid. In our experiments, we tested spheres and hemispheres, for which ζ becomes $\zeta(a, b) = c + (r \cos(a) \sin(b), r \sin(a) \cos(b), r \cos(b))$, where $a \in [0, \pi]$ (or $a \in [0, \pi/2]$ for a hemisphere) and $b \in [0, 2\pi]$. In this context, c and r are constants, representing the sphere’s center and radius.

Second, we introduce an output customization module. Our definition for thematic data is based on a fixed representation of the 3D urban environment. Nonetheless, in urban settings, there is sometimes a need for more abstract metrics to assess the visual experience [25]. For example, the walkability experience can be described as the ratio between pavement and the overall road. We allow the definition of custom perception thematic metrics as algebraic expressions of the thematic vector $\omega : \mathbb{R}^k \rightarrow \mathbb{R}$. The walkability thematic metric can thus be defined as $m_w = \omega_w(\mathbf{m}) = m_{\text{sidewalk}} / (m_{\text{sidewalk}} + m_{\text{road}})$.

Considering the input regularization and output customization modules, the formulation of the general model becomes $F_\Theta = \omega \circ F'_\Theta \circ \rho \circ \zeta$, where F'_Θ is learnable, ρ is fixed, whilst ω, ζ are fixed but adjustable by the user. The generalized output is a customized thematic vector

Algorithm 1 Gradient-Based Optimization for Inverse Query

```

1: Input: Initialized viewpoint  $(x_0, y_0, z_0, \alpha_0, \gamma_0)$ , learning rate  $\eta$ , target thematic vector of view indices  $\mathbf{m}^*$ , tolerance  $\epsilon$ , maximum number of iterations  $N$ 
2: Output: Viewpoint  $(x^*, y^*, z^*, \alpha^*, \gamma^*)$  minimizing  $\|\mathbf{m} - \mathbf{m}^*\|_2^2$ 
3: while  $\|\mathbf{m} - \mathbf{m}^*\|_2^2 > \epsilon$  and iterations  $< N$  do
4:   Compute gradient of loss function:  $\nabla_{x,y,z,\alpha,\gamma} \mathcal{L}$ 
5:   Update viewpoint using gradient descent:
      $(x^*, y^*, z^*, \alpha^*, \gamma^*) \leftarrow (x, y, z, \alpha, \gamma) - \eta \cdot \nabla_{x,y,z,\alpha,\gamma} \mathcal{L}$ 
6:   Evaluate new thematic vector:
      $\mathbf{m} = F_\Theta(x, y, z, \alpha, \gamma)$ 
7: end while
8: Return  $(x^*, y^*, z^*, \alpha^*, \gamma^*)$ 

```

$\mathbf{m} = F_\Theta(a, b) = \omega \circ F'_\Theta \circ \rho \circ \zeta(a, b)$. Notably, all functions that compose F_Θ are differentiable. This allows us to train F'_Θ using standard backpropagation to learn a mapping from viewpoints to the original thematic data. Importantly, the customization modules ζ and ω can be applied at test time without requiring additional training. Given its differentiable nature, we will refer to F_Θ as a neural network for the remainder of this paper.

4.2.3 Training the model

To train F_Θ , we use a dataset of pairs of viewpoints and their corresponding thematic distributions. The ground-truth dataset consists of pairs $\{(x_i, y_i, z_i, \alpha_i, \gamma_i), \mathbf{m}_i^{\text{gt}}\}_{i=1}^N$, where \mathbf{m}_i^{gt} is the vector of thematic value view indices observed from the rasterization of that viewpoint. The ground-truth data will depend on the specific problem and the thematic attributes of interest. For example, in a sunlight access analysis, ground truth may be obtained from sunlight simulation data on the surface of buildings. The data may be evaluated by occlusion from different viewpoints for visibility analysis. This also means that the sampling strategy will depend on the problem. For example, in a sunlight access analysis, viewpoints may be sampled from public open spaces (e.g., parks, streets), while for visibility analysis, viewpoints may be sampled along building surfaces. We will detail specific cases in Section 6.3. The model is trained using the loss function:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \left\| F_\Theta(x_i, y_i, z_i, \alpha_i, \gamma_i) - \mathbf{m}_i^{\text{gt}} \right\|_2^2$$

4.2.4 Querying the model

Once trained, F_Θ provides an efficient way to retrieve thematic distributions for a given viewpoint or find viewpoints satisfying specific conditions. We name these as direct queries (i.e., forward retrieval of thematic data) and inverse queries (i.e., finding desired viewpoints), as shown in Fig. 2.

Direct queries. A direct query will retrieve the thematic distribution \mathbf{m} for a given spatial position and viewing direction, i.e., $\mathbf{m} = F_\Theta(x, y, z, \alpha, \gamma)$. This is a single forward pass through the network. This is particularly efficient given that the model does not require explicit storage of the vector field; it implicitly represents the field as a function. This allows users to quickly extract information about what is visible from a given viewpoint.

Inverse queries. An inverse query will search for inputs $(x, y, z, \alpha, \gamma)$ that minimize the difference between the inferred thematic vector \mathbf{m} and the desired vector of view indices \mathbf{m}^* . This process involves solving:

$$(x^*, y^*, z^*, \alpha^*, \gamma^*) = \arg \min_{(x,y,z,\alpha,\gamma)} \|\mathbf{m} - \mathbf{m}^*\|_2^2$$

We explore two alternatives for performing inverse queries. First, by choosing a large finite subset of viewpoints out of the total space $\mathcal{V}' \subset \mathcal{V}$. The thematic distribution for all the viewpoints can be computed efficiently by adjusting the neural network input batch size accordingly. Out of the computed vectors, the ones closest to the desired vector \mathbf{m}^*

can be returned as approximate solutions. This approach relies on an informed choice for the viewpoint subset \mathcal{V}' .

The second approach for these queries is inspired by the process of inverting neural radiance fields [51]. Given that F_Θ is a neural network, its output \mathbf{m} is differentiable with respect to its input $(x, y, z, \alpha, \gamma)$. We then use gradient-based optimization to iteratively adjust the viewpoint so that the inferred \mathbf{m} becomes as close as possible to the desired target vector \mathbf{m}^* . As a stopping criterion, we use the norm between the inferred thematic vector \mathbf{m} and the desired \mathbf{m}^* , finishing the search when $\|\mathbf{m} - \mathbf{m}^*\|_2^2 \leq \epsilon$, where ϵ is a predefined threshold. Furthermore, we added a maximum number of iterations parameter, N , to allow the user to have better control of the total running time of the algorithm. To explore multiple candidate viewpoints that satisfy the desired thematic conditions, we initialize the optimization process from a set of random locations sampled across the spatial extent of the city. This process is detailed in Algorithm 1.

5 APPLICATIONS IN URBAN ANALYSES

We now illustrate how our model can support view-based exploration of a 3D urban environment. To do so, we use two scenarios to motivate the need for two different tasks.

The first consists of performing visibility analysis on building surfaces (studying *what can be seen from buildings*). This scenario will illustrate the use of direct queries to build visual summaries. The second scenario corresponds to the use of **inverse queries** to find points that have a particular view pattern, given the view distribution of thematic values.

5.1 Direct queries for visibility analysis

Visibility analysis can be performed from different perspectives and at different scales. For example, an architect might want to analyze what can be seen from viewpoints on the surface of many buildings in a neighborhood. The surface of each building is usually segmented into a series of patches that define the resolution for the analysis. This scenario presents two key challenges for interactive visibility data analysis. First, computing visibility-based indices that describe what can be seen from each building surface patch is computationally expensive. As a result, a common approach is to precompute these indices before analysis, using a fixed resolution and a predefined set of buildings. Second, when the number of buildings and the resolution increase, the memory footprint of the resulting visibility data grows significantly. In such cases, memory usage can become difficult to manage, or even prohibitive, limiting both the efficiency and scalability of the analysis.

The model described in Section 4 can be seen as a solution to these challenges. In fact, our model provides a compressive representation of the visibility function with a much smaller footprint than we would need to store the raw data [24]. Furthermore, we can use the model's direct query at arbitrary locations/resolutions within the domain of the training data. Finally, the computation of the direct query is highly parallelizable, and its computation time is independent of the underlying geometry that represents the physical layers of the city.

This makes our model suitable for interactively building a range of customizable visual summaries. For example, we can build heatmap layers to display visibility indices on the surface of buildings. For this, the user can specify a set of target buildings. Given this, we use a customizable resolution parameter to build rectangular patches on the surface of each target building, as shown in Fig. 3(D2). We can use a set of direct queries for each patch to summarize the patch visibility index. For this, we randomly sample a number of viewpoints $P = \{(x_i, y_i, z_i)\}$ within the patch and use the normal vector of the patch to specify the corresponding Euler angles (α, γ) . The thematic value for each patch is computed as $\mathbf{m} = \frac{1}{P} \sum_{i=1}^P F_\Theta(x_i, y_i, z_i, \alpha, \gamma)$.

5.2 Inverse views for pattern-based navigation/exploration

Planning constructions to take advantage of available views is a key concern in architecture [36]. In dense urban environments, views are a scarce and valuable resource. As a result, architects often need to assess which views are accessible from specific areas within the city. Moreover, each building project comes with its own set of constraints,

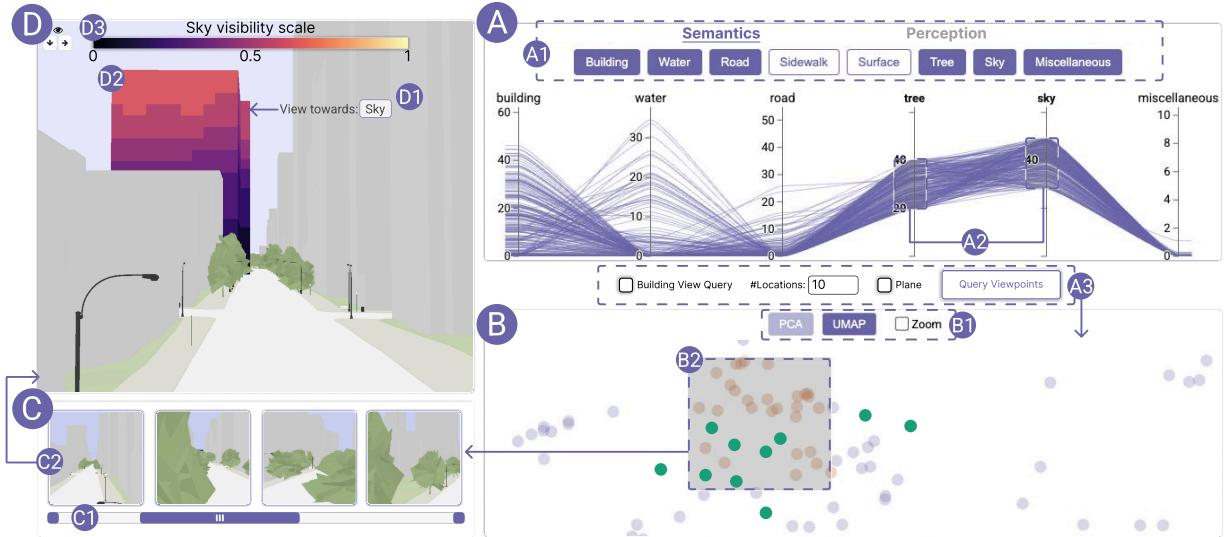


Fig. 3: The interface includes four main components: **A** Control query panel with a PCP for visualizing ground truth visibility distributions. Users can **A1** select distribution type (semantics/perception), **A2** brush axes to define query constraints, and **A3** specify a hyperplane and number of locations to generate new camera views using the model. **B** Latent map view shows model predictions (purple) and generated views (green) in latent space; users can **B1** change projection, zoom/pan, and **B2** brush to highlight points. **C** Gallery view displays the model-generated camera views associated with brushed green points (from **B2**) in **C1** a scrollable panel, and allows **C2** selecting a view to update **D** the 3D map view. The 3D scene supports navigation, with options to **D1** compute visibility of elements like sky, tree, water from building facades, **D2** visualize it on facade tiles, with **D3** a reference color legend.

requiring architects to adapt their search for suitable views based on the project phase and clients' (potentially evolving) preferences. In this context, manually evaluating and exploring viewpoint options can become time-consuming or even infeasible.

Our model can be applied as an automated tool for searching views with desired characteristics in this context. To support this, we offer users three different strategies for conducting such searches. The first approach is designed to broadly explore views across the city using unconstrained gradient-based inverse queries. For instance, a user can search for viewpoints that offer a balanced view between buildings and water by specifying an appropriate view index query vector $\mathbf{m} = (m_b, m_w) = (0.33, 0.33)$. The model optimizes a set of randomly initialized starting locations to approach the specified \mathbf{m} , without imposing any location constraints. As a result, the identified viewpoints may lie in parks, sidewalks, building facades, or other arbitrary locations throughout the city.

The second approach enables users to specify spatial preferences for exploration and desired view composition \mathbf{m} . This type of inverse query is useful for identifying visual highlights along a specific street or within a particular area of the city. To define a target area, the user specifies a plane $\zeta(\cdot, \cdot)$ positioned over the region of interest. Gradient-based optimization is then performed with respect to the plane parameters (a, b) , allowing the model to return locations on the plane – given by $(x, y, z) = \zeta(a, b)$, where $F_\Theta(x, y, z)$ closely matches the target \mathbf{m} .

The third approach for performing inverse queries is tailored to well-structured surfaces that meet two key criteria. First, they have a fixed orientation, meaning the viewing direction is predetermined. Second, they can be divided into tiles or patches with relatively low view variation. A common example of such surfaces in urban environments is building facades. For each facade, we assume a fixed viewing direction aligned with its surface normal, and we partition the facade into rectangular patches. By computing visibility scores for each patch, we can identify and filter those that meet a given view constraint. This method provides an effective way to perform inverse queries for locating desired viewpoints on building facades.

5.3 Visual interface

We have implemented a visual interface that leverages our model to support view-based exploration of 3D urban datasets. The interface,

shown in Fig. 3, is comprised of four components.

Control query panel. The panel (Fig. 3(A)) shows the ground truth visibility distribution using a parallel coordinate plot (PCP) with distribution displayed in terms of view indices (e.g., buildings, sky), or perception (e.g., greenness, walkability) (Fig. 3(A1)). Users can control which variables are displayed in the PCP by using the toggle buttons positioned above the plot (Fig. 3(A)). The PCP is also used for defining query constraints – such as specifying tree visibility between 20–40% and sky visibility between 30–50% through brushing (Fig. 3(A2)), which is then used by our model to estimate new viewpoints that meet these criteria. In addition to brushing, users can specify a 2D hyperplane over parts of the scene and define the number of locations to be retrieved for each query (Fig. 3(A3)). This will prompt the model to estimate viewpoints for the desired number of locations that satisfy the brushing constraints in the hyperplane.

Latent map view. As illustrated in Fig. 3(B), this view presents the model's second-to-last layer predictions projected in a 2D scatter plot using semi-opaque purple points. By default, it displays predictions for the entire ground truth test set; when users brush the PCP (Fig. 3(A2)), the view updates to show only the predictions corresponding to the brushed subset. In addition, the newly estimated viewpoints produced by the model based on the query constraints from the PCP are shown as fully opaque green points in the scatter plot. Users can choose projection type and optionally zoom in and pan the scatter plot (Fig. 3(B1)).

Gallery view. Users can brush the scatter plot to highlight specific points (Fig. 3(B2)), and the corresponding viewpoints for the brushed green points are then displayed in the gallery view (Fig. 3(C)). The gallery view is scrollable (Fig. 3(C1)) and upon selecting one of the viewpoints (Fig. 3(C2)) from the gallery, it will then update the 3D map view with the corresponding camera perspective.

3D map view. The 3D map view (Fig. 3(D)) initially displays the entire 3D urban scene and supports zooming, panning, rotation, and tilt. Upon selecting a viewpoint (Fig. 3(C2)) from the gallery, the 3D scene updates with the corresponding camera perspective. Users can optionally compute the distribution of visibility elements such as sky, tree, or water, from building facades (Fig. 3(D1)) and visualize it over the facade tiles (Fig. 3(D2)). Users can also create queries by brushing the PCP to display specific visibility ranges on facade tiles, which requires enabling the building view query checkbox from the control

query panel. A color legend (Fig. 3(D3)) indicates the distribution of the selected visibility element visualized in the facades.

Design justifications. We chose PCP (Fig. 3(A)) primarily for its ability to create query constraints through simple brushing (Fig. 3(A2)) along multiple axes, making it well-suited for querying multivariate data. Additionally, it effectively visualizes the distribution of ground truth visibility data (e.g., sky, building) in a clear and interpretable manner. To reduce visual clutter when many variables are present, we included toggle buttons (Fig. 3(A1)) that allow users to add or remove axes from the plot as needed. To facilitate exploration of how the ground truth relates to the spatial arrangement of model predictions, we chose to use a scatter plot (Fig. 3(B)) with selectable projection methods (Fig. 3(B1)). Given the high density of data points and potential visual clutter, we enabled zooming and panning (Fig. 3(B1)) for better navigation. To address overlapping points, we reduced their opacity. To distinguish the newly generated camera views resulting from PCP queries, we used a different color (green) for these points and kept them fully opaque, as their number is relatively small. To display the generated viewpoints, we used a scrollable gallery layout (Fig. 3(C1)), as the number of views can often exceed the available window space. We made the 3D scene fully explorable (Fig. 3(D)), supporting pan, zoom, rotate, and tilt interactions. To enable seamless navigation to model-generated viewpoints, we implemented selection of a view from the gallery (Fig. 3(C2)), which automatically updates the 3D scene to that camera perspective. Additionally, to support intuitive visibility exploration of elements (e.g., water, sky) from buildings, we used facade tiles (Fig. 3(D2)) to reveal the distribution of visibility through a color scale (Fig. 3(D3)).

6 EVALUATION

We evaluate our approach through a combination of quantitative and qualitative analyses to assess both the accuracy and performance of the view estimation model, as well as its practical utility in real-world urban analyses. Quantitatively, we measure the accuracy of the predicted view-based distributions against ground-truth data generated via rasterization (Section 6.2.1), and assess computational performance in terms of inference speed (Section 6.2.2). Qualitatively, we demonstrate our approach’s effectiveness across three use cases (Section 6.3) and feedback from experts (Section 6.4).

6.1 Datasets

For all the experiments and use cases, we relied on a publicly available 3D urban dataset with mesh-based representations. In particular, mesh data was downloaded from Geopipe [30]. The dataset includes a region of Manhattan, New York City, where the density of buildings, infrastructure, and vegetation present diverse occlusion and visibility conditions. Our testbed covered 3 km^2 , containing 3,488 building structures and 113,699 vertices.

6.2 Quantitative experiments

6.2.1 Accuracy analysis

Evaluation of prediction accuracy. To evaluate the prediction accuracy of our model, we designed an experiment involving synthetic thematic data, simulating common material-based classes observed in urban environments. We generated six scenarios, each with a different balance between two thematic classes: *brick* and *glass*. For each scenario, buildings in the scene are randomly assigned one of the two classes according to the specified ratio (e.g., 90% brick and 10% glass Fig. 4). This evaluation aimed to test the model’s ability to generalize under varying thematic distributions.

We trained the model for 100 epochs to predict, for each view, the distribution of visible thematic content. Specifically, the percentage of brick versus glass facades observed from that viewpoint. To compute the error, we divided the urban scene into square regions (each roughly corresponding to the size of a city block), and each region was evaluated from 8 uniformly sampled directions. We computed the error between the predicted and ground-truth distributions for each directional view. Table 1 shows the percentage of regions where the prediction error for each class is below 10%. The results show high accuracy across

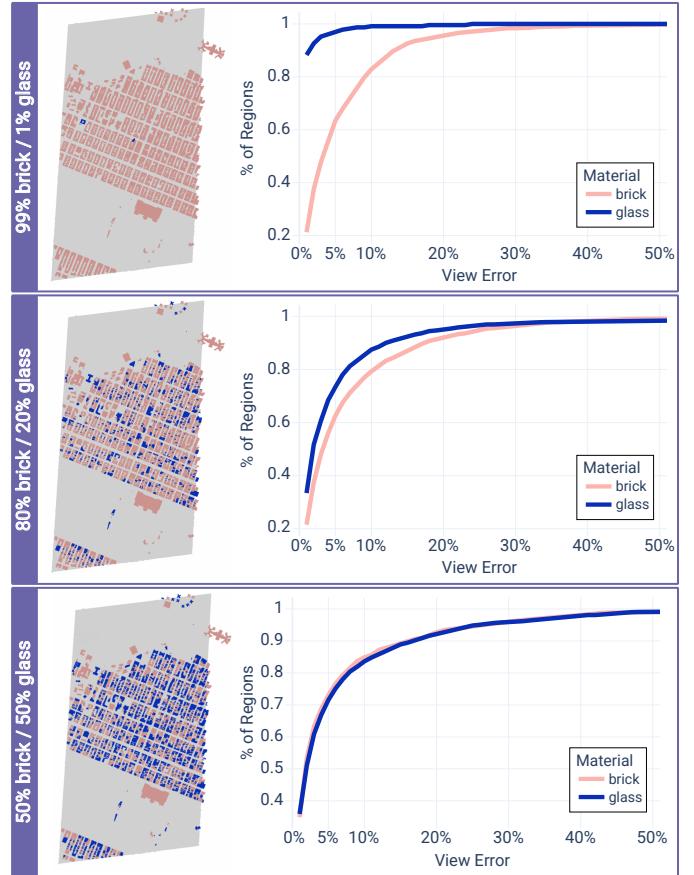


Fig. 4: Three scenarios for the thematic distribution of building materials. On the left side we display aerial views of the scene. On the right we show the cumulative view error of our model with respect to the regions of the scene. In all three scenarios the model predicts the view thematic composition with less than 10% error for $\approx 80\%$ of regions in the scene.

Table 1: Percentage of regions with under 10% error for thematic data with two classes across varying class balance distributions.

Brick / Glass Ratio	99/1	90/10	80/20	70/30	60/40	50/50
Brick regions (%)	82.88	78.26	79.30	81.46	78.24	84.84
Glass regions (%)	99.13	92.43	87.49	84.03	80.83	83.61

a wide range of class balances. For example, in the most imbalanced scenario (99% brick, 1% glass), the model still achieves 82% accuracy on brick regions and 99% on glass regions. Even in the balanced case, performance remains strong, with 84.84% of brick regions and 83.61% of glass regions falling below the 10% error threshold.

Comparison with small-footprint ML models. To evaluate the accuracy of our model relative to other lightweight machine learning models, we conducted a comparative study using a large dataset of labeled views. The task involved predicting the distribution of seven view indices (building, water, road, sidewalk, surface, tree, sky) from the location and view direction of each input. The training set consisted of 63,696 views, and the held-out test set included 15,000 views. To assess performance under different training data densities, we trained our model on subsets of increasing sizes of the full training set (10%, 20%, 30%, 80%, and 100%) and evaluated the root mean squared error (RMSE) on the test set. We conducted the same experiments using two lightweight models: k-nearest neighbors (KNN) with varying numbers of neighbors, as well as a random forest regressor. Each model was trained on identical data splits, and all results were evaluated after 100 epochs.

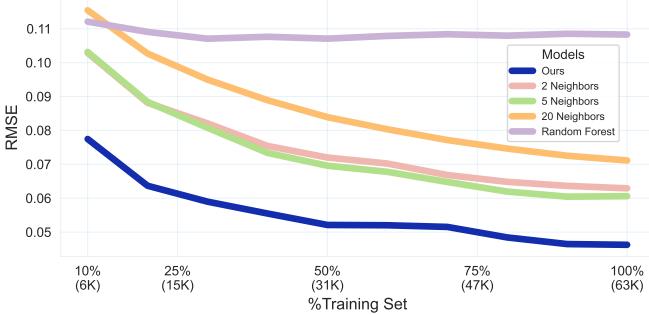


Fig. 5: RMSE of our model compared to baseline small-footprint ML models across varying training set sizes.

As shown in Fig. 5, our model consistently outperforms all other models across training set sizes, particularly in low data scenarios. For example, with only 10% of the training data (6,000), our model achieved an RMSE of 0.078, compared to 0.103 for both the 2- and 5-neighbor KNN models, and 0.112 for the random forest. As the training set size increases, our model continues to improve, achieving 0.046 RMSE on the full dataset.

6.2.2 Computational efficiency analysis

We evaluate the computational performance of our method by measuring the number of views it can infer per second. All performance experiments were conducted on a MacBook Pro M1. Our model achieves an average inference rate of approximately 4,000 views/second, considering a batch size of 1. However, due to the compactness of our model (it only occupies 2.4 MB of memory), much larger batch sizes (exceeding 1,000) are feasible. The effective throughput scales to approximately 4 million views/second in such configurations.

To contextualize these numbers, consider a building with a height of 1,000 feet and a ground-level footprint of 100×100 square feet. Its four facades can be subdivided into 1,000 patches, each measuring 20×20 square feet. Estimating the visibility of each patch from five different viewpoints would require 5,000 view evaluations. Our model can complete this task in approximately 1 to 2 seconds (using a batch size of 1). In contrast, a traditional approach based on rasterization would require over 160 seconds, assuming a rendering rate of 30 frames per second.

6.3 Use cases

We demonstrate the capabilities of our model through a series of use cases, illustrating how urban experts can leverage the view estimation model to explore and analyze view-based thematic information. The first use case highlights how our model can be used for view-based exploration of thematic data. The second use case illustrates how users can define desired view characteristics and use inverse queries to identify viewpoints satisfying custom view constraints. Finally, the third use case focuses on analyzing views from buildings, showing how facade-level viewpoint estimation can support visibility analysis.

6.3.1 Exploration of thematic data using direct queries

In this use case, we demonstrate how our model and system can support the view-based exploration of thematic data. The thematic data is derived from the visibility of buildings towards their surrounding environment. Each building is then annotated with view indices computed from a dense sampling of cameras placed along its facade surfaces.

In a concrete example, an urban analyst might be interested in identifying buildings whose facades have high visibility to trees. The user selects the upper range of the “tree” axis in the query panel, filtering down to views towards buildings that satisfy such a condition. These filtered views are then projected into the 2D latent map, which encodes similarity across views. This view allows users to identify clusters or outliers. Selecting a region in this latent space updates the results gallery, which displays sample views from the selected buildings. These



Fig. 6: Inverse query for views towards **A** trees and sky located on **C** a street. We identify viewpoint clusters using **B** the latent representation of each location generated by our model. In **D**, **E** and **F** we show a selection of views found through the inverse query.

view thumbnails validate the thematic metrics. Clicking a view in the gallery centers the camera on the corresponding viewpoint, allowing users to inspect its geometric surroundings and assess visibility conditions (Fig. 1).

This use case illustrates how our system supports exploration of thematic data without the need for manual 3D navigation. By combining multivariate filtering and the image gallery, users can analyze visibility patterns across the city and ground their interpretations in quantitative data and representative views.

6.3.2 Viewpoint discovery using inverse queries

This use case demonstrates how our model supports *inverse queries* to locate viewpoints that match user-defined goals. Users can specify desired characteristics and allow the system to search for camera positions and directions that yield views satisfying these conditions.

In this case, the user begins by defining the characteristics of interest through the query control panel. In one scenario, they may specify a custom perception metric, built as a weighted combination of values such as openness, visibility of green space, or proximity to landmarks. Alternatively, the user may prioritize high sky exposure and green space visibility while minimizing building density. This could reflect, for instance, a goal to identify locations in the city that feel open, natural, and comfortable for pedestrians [25]. Once a view distribution has been defined, users specify a spatial region over which the query should be applied. This is also done in the query control panel. The user can fine-tune the selected surface directly in the 3D environment if needed.

With the view distribution and spatial domain set, the system initiates the inverse query. It evaluates potential viewpoints within the defined region and returns those whose predicted view distributions most closely match the user’s specification. The top-matching views are presented in the gallery, allowing users to compare candidate perspectives side by side (Fig. 6). This use case illustrates how the system supports goal-driven exploration through inverse queries from desired perceptual conditions.

6.3.3 Quantifying visibility from buildings

This use case focuses on analyzing what is possible to see from the perspective of buildings. Each view is associated with a camera placed on the building’s surface, and is evaluated using the same view indices described in the previous use cases (e.g., sky exposure, water visibility, green space). This enables visibility analysis with applications in real estate [17], architectural evaluation [45], and environmental comfort studies [7].

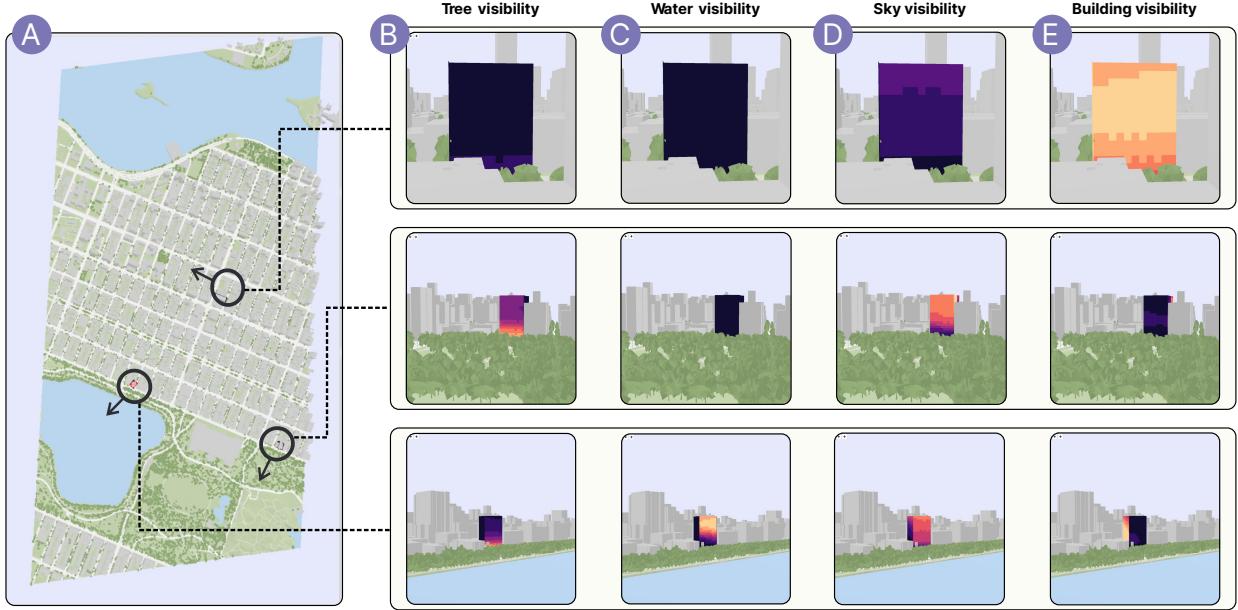


Fig. 7: Comparative direct queries for the thematic data associated with the facades of three buildings. In **A** we show an aerial view of the city with the highlighted locations for each building. In columns **B**, **C**, **D**, and **E** we show the thematic data generated for each facade representing the intensity of views towards trees, water, sky and buildings respectively.

The workflow begins by selecting the building analysis mode in the query control panel, switching the system into a mode where views are computed from building facades rather than towards them. The user then navigates to a building of interest using the 3D exploration map. For example, an analyst might focus on a proposed residential tower near a riverfront. Once the building is selected, the user chooses a view theme to analyze. This might include, for instance, computing how much of the view from each point on the building's facade includes visible water, greenery, or sky. In the query panel, the user can define these thematic goals. The system then evaluates views from multiple sample points on the building's surface, generating a set of view distributions. These are visualized directly on the facade in the 3D component, with colors reflecting the chosen view indices (Fig. 7). Users can switch between different thematic layers to compare how various types of visibility play out across the facade. In parallel, the gallery is populated with individual views from the selected building, sorted by the currently active view index. The gallery allows for a qualitative inspection of the views from the building. Click on any thumbnail links to return to their corresponding position on the facade. This case highlights how the system and model can support building-scale analysis by computing visibility indices, which is especially important to support decisions around design [34, 49], valuation [23], and regulation [4].

6.4 Experts' feedback

To complement our use case demonstrations, we presented the tool to a group of domain experts and collected their impressions. We conducted six 30-minute, semi-structured interviews, during which we explained the goal of our model and demonstrated it through the use cases described in Section 6.3. The experts then reflected on potential applications, limitations, and extensions. We gathered feedback from six experts: E1 is a doctoral student with 3 years of experience in geography and environmental studies. E2 is a postdoctoral researcher with 7 years of experience in design and engineering. E3 is a doctoral candidate with 7 years of experience in design and engineering. E4 is an assistant professor with 12 years of experience in civil engineering. E5 is a doctoral candidate with 15 years of experience in architecture. E6 is an assistant professor with 10 years of experience in urban planning.

Regarding **domain applications**, E1 mentioned that our “method in considering water, trees, and greenery in one composite or thematic version is a breakthrough in visibility analysis.” Regarding applica-

bility for planning, E2 mentioned our method’s usefulness for what-if analyses, mentioning that “Remove the trees, what views are affected? What-if scenarios (...) can be a good asset for planners.” E3 mentioned that “it’s super relevant for urban planning, and it’s something that I’ve never seen before, so I think it’s a really exciting development,” further mentioning that the variables used (open space, water) can “really play an important role in shaping people’s mental health.” Multiple experts also mentioned the value of our contributions for real estate. E2 mentioned that “You could then filter, okay, give me the view that is common for high-income people, or, given an income, what views can I afford?” E3 mentioned that “One of the use cases that could be really relevant would be real estate and finance.” He continued saying that “currently the approaches that we have in real estate are not that good (...) it’s mostly pre-computed and they are not able to simulate new viewpoints in a dynamic way like you have done.”

Regarding **interactivity**, E3 was particularly impressed by the capability of composing indices. He highlighted that “since you have these diverse semantic metrics, you can combine them to find a composite index of how people perceive space,” continuing that “you can simulate many different kinds of indicators that help people understand space better.” E5 stressed that “there is a lot of merit to making these workflows faster (...) making them real-time and interactive to support open-ended exploration.” E5 further highlighted the model’s potential for navigation, mentioning that she could see a use case for finding “the most comfortable route to walk from A to B,” where users choose to maximize “exposure to amenities or open streets or sky view factor, greenery, etc.” She further mentioned that she sees value in not just viewing recommendations but “how can users start viewing it as a way to explore the city, or a route, for instance.”

Experts also highlighted potential directions for **extensions**. E5 highlighted that in the future we could incorporate “additional layers such as urban radiation with watts per square meter so that users could query for the most comfortable spots.” Reflecting on his practice, E1 mentioned that it would be interesting to have users select “landmarks so that views towards them could be analyzed.” E3 mentioned that he would be interested in using our model to create *fingerprints* for buildings, defined as “average of the thematic information of each building,” so that they could be used for “predicting building function.” E6, whose work focuses on sidewalk accessibility, suggested that a valuable extension could be to allow users to “specify the level of

analysis.” In her case, this would mean evaluating views specifically at the street level, with the goal of constructing an index of walkability that incorporates factors such as the visibility of trees and landmarks. Reinforcing E1’s earlier point, E6 also emphasized the importance of enabling users to define landmarks directly within the interface, which she considered a particularly useful addition for walkability-related assessments. Regarding the interface, E6 acknowledged that her first impression was one of “visual overload,” given the number and variety of components presented simultaneously, but found it intuitive after brief exploration. When asked about the latent space, she noted that “its usefulness became clearer through interaction.”

7 CONCLUSION AND DISCUSSION

In this paper, we introduced a view-based 3D urban data exploration technique. At the core of our approach, we developed a neural field model for visibility estimation. We detailed the training pipeline and how to achieve the encoding of physical and thematic data visible from locations throughout the city. We further enhanced the model to include two extensions, one for the input and one for the output. The input regularization module enables a targeted analysis of city subspaces, such as planes hovering over street networks. The output customization module allows the user to define further abstract view indices as algebraic expressions of the default thematic features. As building blocks for data exploration, we proposed two types of operations on top of our visibility estimator: direct and inverse queries. Through direct queries, the user can compute the visibility thematic for a large number of input locations. At the same time, inverse queries allow the user to find spatial locations for which a custom set of visibility constraints is satisfied. We evaluated the performance of our visibility estimator through accuracy tests, comparisons with other small-footprint machine learning models and speed tests against a ground truth scene renderer. As for practical applications, we showcased three urban use cases in which our methodology can be employed. First, we explored thematic data associated with building facades. We showed how to find areas satisfying user-specified visibility constraints on a large number of building facades. Second, we used surface parametrization and gradient-based inverse queries to find views alongside a custom-selected street. Through the exploration of the latent space generated by our model, we were able to get further visibility into insights. Thus, we discovered view sub-patterns within the returned locations. Finally, we estimated views from the perspective of buildings and compared the results based on the location of each building facade. During our meeting with experts, they emphasized our contribution’s relevance across urban planning, real estate, and environmental applications, while also envisioning extensions such as landmark and radiation analyses.

Limitations. We identify a few limitations in our technique and further propose potential solutions. Since our model is trained on a representation of the 3D physical world, there is a natural imbalance between the existing visibility classes. For example, buildings will occupy a much higher percentage of the view space than street furniture. This imbalance leads to regression dilution for the underrepresented classes, in which the model tends to predict lower visibility values for these classes even in areas where they are more present. One solution for this problem would be to use log scales to quantify the thematic data. This scaling would be consistent with the human perceived view impact, which increases logarithmically with respect to the vision stimuli, as mentioned in the *Weber–Fechner Law* [8]. Alternatively, we could use a class-weighted loss, which would dynamically adapt its weights based on the spatial location of the input. Another limitation of our approach is that our model’s precision in visibility could be reduced when modeling larger spacial extents and/or more geometrically detailed areas. One possible solution for this is to train multiple models to cover specific city areas, allowing for higher precision in particular areas. This idea was shown to have a good precision for applications targeting the rendering of urban environments [37]. Finally, our model assumes mesh-based urban representations, which may limit its direct applicability to other formats such as point clouds or textured models.

Future work. We identified several opportunities for extending our current work. First, while our approach is in principle agnostic to

the specific urban or suburban setting, our evaluation has focused on environments where building geometry plays a central role. Extending it to less dense and rural settings (e.g., landscapes with minimal built structures) is an interesting avenue for future work, particularly for applications in landscape analysis [42] and rural spatial planning [15]. We also plan to integrate a number of additional types of thematic data into our analysis, such as sunlight access, flooding exposure, noise propagation, and traffic intensity. Second, we plan to implement this technique as part of our growing open source toolkit for urban data visualization [31, 32]. Third, a direct path towards improving the precision of our model would be dividing the scene into regions and training individual specialized models for each area. Furthermore, the model could be extended to support “what-if” scenarios, for instance, by simulating the view impact of a new construction through regenerating training data in its proximity and fine-tuning the model on the fly with this augmented dataset. Finally, a promising direction for future work is a dedicated study with urban experts to evaluate how latent space representations of views can support domain-specific analysis and decision-making.

ACKNOWLEDGMENTS

We thank the experts who participated in our evaluation and the reviewers for their constructive feedback. This work was supported by the U.S. National Science Foundation (Awards #2320261, #2330565, #2411223), IDOT (TS-22-340), CNPq (311425/2023-2) and conducted as part of the Open-Source Cyberinfrastructure for Urban Computing (OSCUR) project.

REFERENCES

- [1] M. Alfakhor, H. Dastageeri, S. Schneider, and V. Coors. Occlusion screening using 3D city models as a reference database for mobile ar-applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W3-2022:11–18, 2022. 1
- [2] X. Bonaventura, M. Feixas, M. Sbert, L. Chuang, and C. Wallraven. A survey of viewpoint selection methods for polygonal models. *Entropy*, 20(5), 2018. 3
- [3] S. Boorboor, Y. Kim, P. Hu, J. M. Moses, B. A. Colle, and A. E. Kaufman. Submerse: Visualizing storm surge flooding simulations in immersive display ecologies. *IEEE Transactions on Visualization and Computer Graphics*, 30(9):6365–6377, 2024. 1, 2, 3
- [4] M. Brasebin, J. Perret, S. Mustière, and C. Weber. 3D urban data to assess local urban regulation influence. *Computers, Environment and Urban Systems*, 68:37–52, 2018. 9
- [5] S. Chen, F. Miranda, N. Ferreira, M. Lage, H. Doraiswamy, C. Brenner, C. Defanti, M. Koutsoubis, L. Wilson, K. Perlin, and C. Silva. UrbanRama: Navigating cities in virtual reality. *IEEE Transactions on Visualization & Computer Graphics*, 28(12):4685–4699, Dec. 2022. 1
- [6] Z. Chen, H. Qu, and Y. Wu. Immersive urban analytics through exploded views. In *Workshop on Immersive Analytics*, 2017. 3
- [7] F. De Luca. Sun and wind. integrated environmental performance analysis for building and pedestrian comfort. *Proc. of the 2019 Symposium on Simulation for Architecture and Urban Design (San Diego, USA, 2019), SimAUD ’19*, p. 1–8, 02 2019. 8
- [8] S. Dehaene. The neural basis of the weber–fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4):145–147, 2003. 10
- [9] H. Deng, L. Zhang, X. Mao, and H. Qu. Interactive urban context-aware visualization via multiple disocclusion operators. *IEEE Trans. on Visualization and Computer Graphics*, 22(7):1862–1874, 2016. 2
- [10] H. Doraiswamy, N. Ferreira, M. Lage, H. Vo, L. Wilson, H. Werner, M. Park, and C. Silva. Topology-based catalogue exploration framework for identifying view-enhanced tower designs. *ACM Trans. Graph.*, 34(6), article no. 230, 13 pages, Nov. 2015. 1, 2, 3
- [11] N. Elmquist and P. Tsigas. A taxonomy of 3D occlusion management for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1095–1109, 2008. 2
- [12] N. Ferreira, M. Lage, H. Doraiswamy, H. Vo, L. Wilson, H. Werner, M. Park, and C. Silva. Urbane: A 3D framework to support data driven decision making in urban development. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 97–104, 2015. 1, 2, 3

- [13] P. Florio, G. Peronato, A. Perera, A. Di Blasi, K. H. Poon, and J. H. Kämpf. Designing and assessing solar energy neighborhoods from visual impact. *Sustainable Cities and Society*, 71:102959, 2021. 3
- [14] C. R. Fol, N. Shi, N. Overney, A. Murtiyoso, F. Remondino, and V. C. Griess. 3D dataset generation using virtual reality for forest biodiversity. *International Journal of Digital Earth*, 17(1):2422984, 2024. 3
- [15] K. Fujiwara, R. Tsurumi, T. Kiyono, Z. Fan, X. Liang, B. Lei, W. Yap, K. Ito, and F. Biljecki. Voxcity: A seamless framework for open geospatial data integration, grid-based semantic 3D city model generation, and urban environment simulation. *arXiv preprint arXiv:2504.13934*, 2025. 10
- [16] K. Genova, M. Savva, A. X. Chang, and T. A. Funkhouser. Learning Where to Look: Data-driven viewpoint set selection for 3D scenes. *CoRR*, abs/1704.02393, 2017. 3
- [17] D. Kaimaris, A. Xifilidou, and N. Karanikolas. Visibility and real estate values. A room with view is a room with value? *Journal of Engineering and Architecture*, 01 2017. 8
- [18] D. Kasraian, S. Adhikari, D. Kossowsky, M. Luubert, G. B. Hall, J. Hawkins, K. N. Habib, and M. J. Roorda. Evaluating pedestrian perceptions of street design with a 3D stated preference survey. *Environment and Planning B: Urban Analytics and City Science*, 48(7):1787–1805, 2021. 1
- [19] B. Lei, P. Janssen, J. Stoter, and F. Biljecki. Challenges of urban digital twins: A systematic review and a delphi expert survey. *Automation in Construction*, 147:104716, 2023. 1
- [20] J. Li, Z. Xiao, and J. Kong. A viewpoint based approach to the visual exploration of trajectory. *Journal of Visual Languages & Computing*, 41:41–53, 2017. 3
- [21] M. Li, F. Xue, Y. Wu, and A. Yeh. A room with a view: Automatic assessment of window views for high-rise high-density areas using city information models and deep transfer learning. *Landscape and Urban Planning*, 226:104505, 2022. 2
- [22] Q. Li, Y. J. Liu, L. Chen, X. C. Yang, Y. Peng, X. R. Yuan, and M. L. L. Wijerathne. SEEVis: A smart emergency evacuation plan visualization system with data-driven shot designs. *Computer Graphics Forum*, 39(3):523–535, 2020. 3
- [23] J. Lu. The value of a south-facing orientation: A hedonic pricing analysis of the Shanghai housing market. *Habitat International*, 81:24–32, 2018. 9
- [24] Y. Lu, K. Jiang, J. A. Levine, and M. Berger. Compressive neural representations of volumetric scalar fields. In *Computer Graphics Forum*, vol. 40, pp. 135–146, 2021. 4, 5
- [25] X. Ma, C. Ma, C. Wu, Y. Xi, R. Yang, N. Peng, C. Zhang, and F. Ren. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. *Cities*, 110:103086, 2021. 4, 8
- [26] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 4
- [27] F. Miranda, H. Doraiswamy, M. Lage, L. Wilson, M. Hsieh, and C. T. Silva. Shadow Accrual Maps: Efficient accumulation of city-scale shadows over time. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1559–1574, 2019. 1, 3
- [28] F. Miranda, T. Ortner, G. Moreira, M. Hosseini, M. Vuckovic, F. Biljecki, C. T. Silva, M. Lage, and N. Ferreira. The state of the art in visual analytics for 3D urban data. *Computer Graphics Forum*, 43(3), 2024. 1, 2, 3
- [29] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, Y. Bai, and A. Behnood. 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Advanced Engineering Informatics*, 51:101501, 2022. 3
- [30] C. Mitchell and T. D. Dickerson. Geopipe developer portal. <https://docs.geopipe.ai>. 7
- [31] G. Moreira, M. Hosseini, M. N. Alam Nipu, M. Lage, N. Ferreira, and F. Miranda. The Urban Toolkit: A grammar-based framework for urban visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1402–1412, 2024. 10
- [32] G. Moreira, M. Hosseini, C. Veiga, L. Alexandre, N. Colaninno, D. de Oliveira, N. Ferreira, M. Lage, and F. Miranda. Curio: A dataflow-based framework for collaborative urban visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1224–1234, 2025. 10
- [33] R. Mota, N. Ferreira, J. D. Silva, M. Horga, M. Lage, L. Ceferino, U. Alim, E. Sharlin, and F. Miranda. A comparison of spatiotemporal visualizations for 3D urban analytics. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1277–1287, 2023. 2
- [34] J. L. Nasar. Urban design aesthetics: The evaluative qualities of building exteriors. *Environment and Behavior*, 26(3):377–401, 1994. 9
- [35] R. Neuville, J. Pouliot, F. Poux, and R. Billen. 3D viewpoint management and navigation in urban planning: Application to the exploratory phase. *Remote Sensing*, 11(3), 2019. 3
- [36] T. Ortner, J. Sorger, H. Steinlechner, G. Hesina, H. Piringer, and E. Gröller. Vis-A-Ware: Integrating spatial and non-spatial visualization for visibility-aware urban planning. *IEEE Transactions on Visualization and Computer Graphics*, 23(2):1139–1151, 2017. 1, 2, 5
- [37] M. Partha, S. Gupta, and G. Gao. Neural city maps: A case for 3D urban environment representations based on radiance fields. In *Proceedings of the 36th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2023)*, pp. 1953–1973, 2023. 10
- [38] M. Quintana, Y. Gu, and F. Biljecki. My street is better than your street: Towards data-driven urban planning with visual perception. In *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 221–222, 2024. 1
- [39] M. Ranzinger and G. Gleixner. GIS datasets for 3D urban planning. *Computers, Environment and Urban Systems*, 21(2):159–173, 1997. Urban Data Management Symposium. 3
- [40] S. Salimi, M. Mirgholami, and A. Shakibamanesh. Visual impact assessment of urban developments around heritage landmarks using ULVIA method. *Environment and Planning B: Urban Analytics and City Science*, 50(3):678–693, 2023. 1, 2
- [41] Y. Sun, Q. Huang, D.-Y. Hsiao, L. Guan, and G. Hua. Learning view selection for 3D scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14464–14473, June 2021. 3
- [42] A. R. Swietek and M. Zumwald. Visual capital: Evaluating building-level visual landscape quality at scale. *Landscape and Urban Planning*, 240:104880, 2023. 10
- [43] N. Tarkhan, J. T. Szcześniak, and C. Reinhart. Façade feature extraction for urban performance assessments: Evaluating algorithm applicability across diverse building morphologies. *Sustainable Cities and Society*, 105:105280, 2024. 1
- [44] I. Turan, A. Chegut, D. Fink, and C. Reinhart. Development of view potential metrics and the financial impact of views on office rents. *Landscape and Urban Planning*, 215:104193, 2021. 2
- [45] T. Varoudis and P. Sophia. Beyond two dimensions architecture through three dimensional visibility graph analysis. *The Journal of Space Syntax* 5, pp. 91–108, 2014. 8
- [46] J.-P. Virtanen, K. Jaalama, T. Puustinen, A. Julin, J. Hyypä, and H. Hyypä. Near real-time semantic view analysis of 3D city models in web browser. *ISPRS International Journal of Geo-Information*, 10(3):138, 2021. 1, 2
- [47] B. Willenborg, M. Sindram, and T. H. Kolbe. Applications of 3D city models for a better understanding of the built environment. *Trends in Spatial Analysis and Modelling: Decision-Support and Planning Strategies*, pp. 167–191, 2018. 3
- [48] P. P.-J. Yang, S. Y. Putra, and W. Li. Viewsphere: A GIS-based 3D visibility analysis for urban design evaluation. *Environment and Planning B: Planning and Design*, 34(6):971–992, 2007. 3
- [49] W. Yap, J.-H. Chang, and F. Biljecki. Incorporating networks in semantic understanding of streetscapes: Contextualising active mobility decisions. *Environment and Planning B: Urban Analytics and City Science*, 50:1416–1437, 2023. 9
- [50] S. Yasumoto, A. Jones, T. Nakaya, and K. Yano. The use of a virtual city model for assessing equity in access to views. *Computers, Environment and Urban Systems*, 35(6):464–473, 2011. 1, 2
- [51] L. Yen-Chen, P. Florene, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin. INeRF: Inverting neural radiance fields for pose estimation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1323–1330, 2021. 5
- [52] S. Yu, B. Yu, W. Song, B. Wu, J. Zhou, Y. Huang, J. Wu, F. Zhao, and W. Mao. View-based greenery: A three-dimensional assessment of city buildings' green visibility using Floor Green View Index. *Landscape and Urban Planning*, 152:13–26, 2016. 1, 2
- [53] C. Zhang, W. Zeng, and L. Liu. UrbanVR: An immersive analytics system for context-aware urban design. *Computers & Graphics*, 99:128–138, 2021. 3
- [54] H. Zhang, S. Nijhuis, and C. Newton. Uncovering the visibility of blue spaces: Design-oriented methods for analysing water elements and maximizing their potential. *Journal of Digital Landscape Architecture*, 2023(8):628–638, 2023. 1, 2