Privacy Leakage from a Thousand Words: Millipixel Location Recovery from Dot Maps

Anonymous author(s)

***Abstract*—Dot maps, which visualize individual data points as dots over a geographic region, are widely used across scientific papers and articles to reveal spatial patterns and relationships. However, the understanding of the privacy risks associated with dot maps that cover large geographic areas remains limited. In this paper, we systematically analyze these risks and present AutoLocate, an automated framework for high-precision location recovery. At its core, AutoLocate exploits the anti-aliasing techniques applied by default during map generation. We develop two complementary recovery algorithms that use these anti-aliasing artifacts to reverse-engineer dot locations, tailored to scenarios where the adversary either has or lacks access to the map background. Extensive experiments across a wide range of map scales, resolutions, dot styles, and visualization platforms demonstrate the effectiveness of AutoLocate. In particular, it achieves average recovery errors as low as 1 meter (approximately 0.0005 pixel precision) on maps covering large regions (e.g., the United States), which is over 200 times more accurate than existing approaches. Given the severity of this threat, we also propose mitigation strategies and introduce a privacy risk assessment tool to help researchers evaluate and reduce privacy leakage when publishing dot maps.**

1. Introduction

A well-known adage in communication is "A picture is worth a thousand words". A commonly used type of picture is the dot map (also known as the dot distribution/density map), which employs point symbols to visualize the geographic distribution of a large number of related phenomena. Dot maps rely on visual scatter to show spatial patterns, especially variations in density. They are often used in important fields such as medical research, urban planning, and environmental studies [1], [2]. By representing each instance's location as a "dot" (which may take the form of a circle, triangle, or other symbols) on a map, researchers can detect spatial patterns, identify clusters, and trace potential sources of outbreaks [3]–[5]. For instance, Soetens et al., [4] demonstrate the use of dot maps in Germany and the Netherlands by plotting individual disease cases to reveal their spatial distribution and highlight outbreak clusters. The rapid development of map visualization platform, ranging from professional Geographic Information Systems (GIS) (*e.g.,* ArcGIS [6] and QGIS [7]) to commercial visualization tools (*e.g.,* Tableau [8]) and programming libraries (*e.g.,* GeoPandas [9] and R [10]), has made it easy for researchers to obtain precise geolocation information and publish highly accurate dot maps.

While visualizations from dot maps offer clear insights into spatial relationships, they also raise significant concerns about the privacy of individuals represented on the map. This issue becomes particularly critical in privacy-sensitive domains such as healthcare, where dot maps are used for disease surveillance, risk assessment, and monitoring of public health trends [1], [11]. As highlighted in previous studies [12], the publication of raw geospatial data can introduce serious risks, including threats to personal safety from targeted crimes, legal and ethical violations due to privacy breaches, and social consequences such as neighborhood stigmatization. Despite these risks, dot maps remain a widely adopted and indispensable spatial visualization and analysis tool, with their use continuing to grow across a broad range of disciplines (see Section 2 for a detailed overview). Therefore, to balance individual privacy and utility, it is essential to develop methods that can accurately assess privacy risks when publishing dot maps.

Prior studies have investigated these risks by examining how accurately locations can be recovered from dot maps [12]–[14]. In these works, researchers first identify each dot's centroid using methods such as manual visual inspection [13], [15]–[17] or unsupervised learning [18]. They then encode the estimated centroid to its corresponding geographic coordinates and use the resulting recovery error as a measure of privacy risk. Using these approaches, several studies [17], [19] have demonstrated that it is possible to recover individual locations from dot maps at the neighborhood or city level, with average errors around 100 meters. For instance, one study [16] found that patient locations could be re-identified with an average error of 96.38 meters from a dot map of a parish in the United States.

To the best of our knowledge, there has been little progress in the past two decades on location recovery from dot maps. Specifically, we are unaware of any research that has investigated the privacy risks associated with dot maps representing large geographic areas, such as countries or continents. In such maps, each pixel represents a larger area, and existing approaches can only recover locations with substantial errors, often on the order of hundreds of meters. As a result, researchers may implicitly assume that publishing maps covering large geographic areas presents limited privacy concerns. Moreover, there is no systematic framework for assessing the risks, and a comprehensive understanding of how map properties (*e.g.,* scale, resolution,

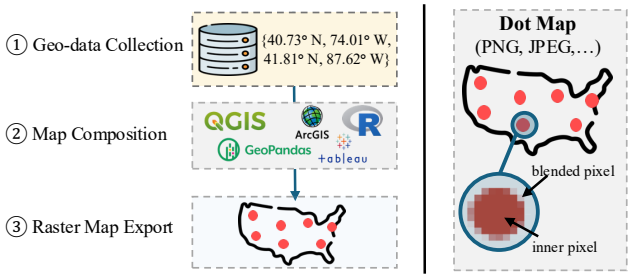
and dot styles) impact privacy leakage remains elusive.

In this paper, we address these gaps by conducting a systematic analysis of privacy risks in publishing dot maps. Our key insight is that location recovery accuracy can be dramatically improved by exploiting the *anti-aliasing artifacts* used to render dot symbols when generating maps. Anti-aliasing [20], [21] is a standard rendering technique that smooths jagged edges by blending the colors of boundary pixels of objects according to how much of each pixel is covered by the underlying shape (demonstrated in Figure 1(b)). While this mechanism improves map visual quality, the resulting blended colors inadvertently encode sub-pixel information about the dot's precise geometric centroid. By reverse-engineering these artifacts, it is possible to recover a dot's coordinates with a precision that far surpasses simple pixel centroid detection or visual inspections.

Building on this insight, we introduce AutoLocate, an automated location recovery framework that leverages anti-aliasing to infer precise geographic coordinates from dot maps. The framework operates under two attack scenarios, depending on whether the adversary has access to the background map. When the background map is available, AutoLocate utilizes a perceptual boundary alignment algorithm that iteratively refines a dot's centroid location to better match the target dot's anti-aliased boundary. When the background is unknown, it applies a geometric refinement algorithm that models the geometric characteristics of dot symbols and compares the expected and observed boundary color patterns to accurately infer the dots' centroids. Extensive experiments on several popular map visualization tools (*i.e.,* QGIS, GeoPandas, and R) demonstrate the effectiveness and robustness of AutoLocate across a wide range of map scales, backgrounds, resolutions, formats, and dot distributions. In particular, our experiments show that AutoLocate dramatically improves location recovery accuracy, achieving average errors of **approximately 1 meter (0.0005 pixel precision)** on maps covering large regions (*e.g.,* maps of the United States), outperforming existing methods by up to $\mathbf{200}\times$ in recovery accuracy.

Our work challenges a hidden assumption in geographic data visualization that scale alone protects privacy, and it highlights the importance of examining how these maps are constructed when assessing their privacy risks. Given the severity of this threat, we also explore several mitigation strategies and develop a privacy risk assessment tool. This tool integrates our framework with underlying population density to help researchers evaluate and mitigate privacy leakage when publishing dot maps. In summary, we make the following contributions:

- We systematically study the privacy risks of dot maps by proposing an automated location recovery framework named AutoLocate.
- We design location recovery algorithms that exploit anti-aliasing artifacts of dot maps for precise location estimation, depending on whether the map background is available to the adversary.
- Extensive experiments demonstrate that AutoLocate is highly effective, achieving over 200 times lower error at recovering dot locations than prior approaches, and remains robust across different types of dot maps.
- We introduce a risk assessment tool based on AutoLocate to help researchers evaluate and mitigate the privacy risks of their maps.



(a) Map generation process (b) Anti-aliasing

Figure 1: (a) Illustration of the map generation process: high-precision geographic coordinates of individuals are collected, followed by map composition using visualization software, and then exported to a raster image format (*e.g.,* PNG, JPEG, *etc.*) for publication or dissemination. (b) A close-up of the resulting raster dot map, illustrating the anti-aliasing artifacts (*i.e.,* blended pixels) at the dot's boundary.

The rest of this paper is organized as follows. Section 2 introduces the necessary background on dot maps and the map generation process. Section 3 defines the threat model and attack scenarios. We then detail our proposed location recovery framework (AutoLocate) in Section 4. Section 5 presents the experimental results of the proposed attacks. Section 6 discusses mitigation strategies and the proposed privacy risk assessment tools. Related work is detailed in Section 7, and the paper concludes in Section 8.

2. Background

Dot Maps Usage in Academic Research. Dot maps have a long history as an important tool for scientific discovery. One of the most famous early examples is John Snow's 1854 Broad Street cholera map [22], which plotted individual cholera cases as dots, enabling the visual identification of a contaminated water pump as the source of the outbreak. This seminal work is widely regarded as a foundational event in modern epidemiology, demonstrating the power of spatial visualization for public health analysis. Since then, dot maps have become an essential visualization method across a wide range of research domains, including public health, urban planning, environmental monitoring, disaster management, and socioeconomic analysis. To demonstrate their prevalence, we conducted a literature survey using the keywords such as "dot map" (see detailed survey process in Appendix A). Table 1 summarizes exemplary publications that employed dot maps across these domains. As shown, researchers frequently employ dot maps to visualize sensitive information such as patient home addresses, crime incident locations, and household demographics. This practice spans various map scales, ranging from city-level to national-level

TABLE 1: Examples of dot maps used for visualizing and analyzing sensitive data in academic research.

Research Field	Example Sensitive Data	Publications
Public Health	Patient home addresses	[23]–[26]
Criminology	Crime incident locations	[27], [28]
Urban Planning	Home–work locations	[29], [30]
Ecology	Endangered species locations	[31]–[34]
Social Science	Household demographics	[35]–[37]
Education	Student residences	[38], [39]
Archaeology	Artifact find spots	[40], [41]

visualizations. The widespread adoption of dot maps for sensitive data visualizations highlights the need to systematically assess the privacy risks associated with these maps.

We also observe that many of these published dot maps contain, and are often dominated by, isolated dots. In addition, most employ circular dot symbols. As shown in our experiments in Section 5, our attacks achieve high recovery precision on such isolated circular dots, revealing significant privacy implications for this common visualization practice.

Dot Map Generation Process. The generation of a dot map typically involves three main phases, as illustrated Figure 1(a). (i) *Data Collection.* In this phase, high-precision geographic coordinates of individual subjects (*e.g.,* patient home addresses) are obtained. This could be done through various methods, such as GPS devices, address geocoding, or extracting data from existing datasets. (ii) *Map Composition.* The collected geospatial data are then imported into map visualization software, where they are internally represented as vector graphics, which use mathematical primitives like points and polygons that are resolution-independent [42], [43]. Each data point is displayed as a dot on the map, layered over a background that may include country borders, satellite imagery, or a simple white canvas. Key processing steps in this phase include map projection, background selection, and visual encoding choices (*e.g.,* dot type, size, and color) to enhance visibility and readability. (iii) *Map Export.* Once the map is finalized, the combination of the dot layer and the background layer is exported to a raster image format (*e.g.,* PNG and JPEG) for publication or dissemination. While the coordinates of dots can be directly obtained from vector graphics, recovering high-precision locations from the final rasterized map is a non-trivial task.

Rasterization and Anti-Aliasing. The Map Export phase relies on *rasterization*, a process that converts resolution-independent vector graphics into a discrete grid of pixels. To improve image quality during rasterization, spatial anti-aliasing [20], [21] is applied **by default**. Specifically, anti-aliasing is a computer graphics technique that is used to smooth the edges of graphical elements, reducing distortion artifacts (*i.e.,* aliasing) that occur when representing lossless vector data (or high-resolution image) at a lower pixel resolution [44]. The core principle of anti-aliasing is to add intermediate shades of color to pixels along the edges of objects based on the proportion of each vector edge that covers a pixel [45], [46]. This blending softens object boundaries and enhances visual smoothness, reducing the harsh, pixelated appearance and making lines and curves appear smoother and more natural (as demonstrated in Figure 1(b)). Over the past decades, various anti-aliasing methods have been developed to balance visualization quality and rendering efficiency, such as Supersampling Anti-Aliasing (SSAA) [47], Multisampling Anti-Aliasing (MSAA) [48], and Fast Approximate Anti-Aliasing (FXAA) [49].

We observe that the subtle color gradients produced by anti-aliasing along the edges of rendered dots can be exploited to infer the centroid of each dot with higher accuracy, potentially down to the millipixel level.

3. Threat Model and Attack Scenarios

In this section, we outline the threat model and attack scenarios considered in this paper. Specifically, we focus on the problem of location recovery in dot maps, where an adversary tries to reverse-engineer the geographic coordinates of individual dots from a raster map image.

Adversary's Goal. Given a raster map image $\mathbf{I} \in \mathbb{R}^{W \times H \times 3}$ (with width W, height H, and RGB color channels), where each dot represents an individual's location, the adversary's goal is to estimate the true geographic coordinates (*i.e.,* latitude and longitude) corresponding to each dot.

Advesary's Capbilities. We make the following realistic assumptions about the adversary's capabilities:

- *Dot Properties.* The adversary knows the visual properties of the target dots, including their geometry ϕ (*e.g.,* circle, triangle, or pentagon), size ρ, and color $\mathbf{z}$. These properties can be easily determined using visual inspection and image editing tools (*e.g.,* a pixel selector).
- *Coordinate Transformation.* The adversary can learn a coordinate transformation function, *i.e.,* $\mathcal{F} : (x, y) \mapsto (\text{lat}, \text{lon})$, which maps coordinates (x, y) in the raster map to geographic coordinates. This transformation function can be derived from map legends (which provide scale and projection details) or reconstructed using the georeferencing features of modern GIS tools (*e.g.,* ArcGIS, QGIS) to align the map with a known coordinate system.

Baseline: PixelMatch. A simple method to estimate the location of a dot involves calculating the centroid of the pixels that the dot occupies, which we call PixelMatch. As shown in Algorithm 1, the algorithm first identifies all pixel clusters that match the dot's color using a standard Breadth-First Search (BFS) algorithm (line 1, `FindPixelClusters` function). For each pixel cluster, it computes the mean coordinates of all the pixels in the cluster (line 4). Finally, these pixel centroids are transformed into geographic coordinates using the function $\mathcal{F}$ to obtain the final estimated location.

Missed Opportunities of Existing Approaches. Previous studies [15], [16], [18] have used the idea of PixelMatch, either through manual visual inspection or by using GIS tools to estimate the centroids of dots to recover locations. However, these approaches fail to fully exploit the available information and do not account for realistic scenarios in which dots may overlap:

Algorithm 1 Baseline: PixelMatch. It identifies connected pixels with the same color as the target dot and computes their mean coordinate as the estimated location.

Require: Target dot map $\mathbf{I}$, dot color $\mathbf{z}$
$\mathcal{P} \leftarrow \texttt{FindPixelClusters}(\mathbf{I}, \mathbf{z})$
$\mathcal{C} \leftarrow \{\}$ ▷ *initialize estimated locations*
for each cluster $\mathbf{P} \in \mathcal{P}$ **do**
  $(x_c, y_c) \leftarrow \frac{1}{|\mathbf{P}|} \sum_{(x,y)\in\mathbf{P}} (x, y)$
  $\mathcal{C} \leftarrow \mathcal{C} \cup \{x_c, y_c\}$
end for
return $\mathcal{C}$

- *Anti-aliasing Artifacts on Dot Boundaries.* During the rasterization process, anti-aliasing blends the boundary pixels of dots with the background. The color of these boundary pixels is sensitive to the location of the dot's centroid and could be leveraged to improve location recovery. However, existing approaches overlook these boundary pixels because their color does not exactly match the dot color. Moreover, as demonstrated in Section 5, simply incorporating boundary pixels for centroid estimation is still ineffective for high-precision location recovery.
- *Overlapping Dots.* In some cases, dots on the map may overlap, forming dot clusters where the boundaries of individual dots become unclear. This overlap makes it more difficult to recover the precise location of each dot. Existing methods fail to account for this scenario, leaving the challenge of overlapping dots largely unaddressed.

Attack Scenarios. These limitations motivate us to develop a systematic framework for location recovery that better leverages anti-aliasing artifacts and handles dot overlaps. Since the anti-aliasing boundary pixels are blended with both the dot and its neighboring background, having knowledge of the background could help in more effectively utilizing these artifacts. Thus, we consider two attack scenarios based on the availability of the map's background:

- *Background Known.* The adversary has access to the background $\mathbf{B}$ used to generate the dot map $\mathbf{I}$. This is often the case even when the adversary does not have access to the original background. Many backgrounds in dot maps from publications (as shown in Table 1) are simple uniform colors (*e.g.,* white) or sourced from publicly available repositories (*e.g.,* OpenStreetMap [50]), making them easy to replicate. Additionally, we assume the adversary has access to a visualization tool (*e.g.,* a map render function $\mathcal{R}$) that is the same as, or similar to, the one used to generate the target map, enabling them to simulate its map generation process.
- *Background Unknown.* The adversary only has access to the final map $\mathbf{I}$, without the background $\mathbf{B}$. In this case, the adversary must infer the locations of the dots solely from the rasterized map, without access to the background or visualization tools.

It is worth mentioning that, while anti-aliasing is enabled by default during map generation, our attack does not require knowledge of the specific anti-aliasing technique used. Instead, we treat the map rendering function $\mathcal{R}$ as a black box. Furthermore, although dot maps can be exported in various formats that may include additional information (*e.g.,* PNG's alpha channel for transparency or metadata in TIFF), our attack does not rely on this extra information. This ensures that our approach remains effective across all common image formats, as demonstrated in Section 5.2.

4. AutoLocate: A Framework for Automated Location Recovery from Dot Maps

In this section, we present AutoLocate, an automated, high-precision framework for location recovery in dot maps. Specifically, we introduce two recovery algorithms: one designed for an adversary with knowledge of the target map's background (*i.e.,* perceptual boundary alignment), and the other for an adversary without access to the background (*i.e.,* geometric refinement).

4.1. Perceptual Boundary Alignment (PBA)

We first consider the attack scenario where the adversary possesses the background $\mathbf{B}$ used to compose the target dot map. In this setting, the adversary can generate new dot maps using a set of estimated dot coordinates and then compare these maps with the target map. By analyzing the differences at *anti-aliased boundaries*, the adversary iteratively refines the coordinates until the rendered map closely matches the target. This transforms the location recovery problem into an optimization task: find the geometric centroid of each dot that minimizes the visual mismatch between the rendered and target maps. We first define the optimization loss function, detail how to compute and optimize it, and then present the full recovery algorithm.

Perceptual Loss Function. Given a set of estimated dot centroids $\mathcal{C}$, the dot's properties (color $\mathbf{z}$, shape ϕ, size ρ), and the background $\mathbf{B}$, the adversary generates a new dot map $\mathbf{I}'$ using a map rendering function $\mathcal{R}$:

$$\mathbf{I}' \leftarrow \mathcal{R}(\mathbf{B}, \mathcal{C}, \mathbf{z}, \phi, \rho).$$

A good location estimate should produce a rendered map that perceptually matches the target map $\mathbf{I}$, particularly around anti-aliased boundary pixels where subtle dot centroid shifts produce measurable color changes. For a pixel at location (x, y), we define the perceptual discrepancy function $d(\cdot, \cdot)$ using the L_1 distance between RGB channels of the two maps:

$$d\left(\mathbf{I}_{x,y}, \mathbf{I}'_{x,y}\right) = \left|\mathbf{I}^r_{x,y} - \mathbf{I}'^r_{x,y}\right| + \left|\mathbf{I}^g_{x,y} - \mathbf{I}'^g_{x,y}\right| + \left|\mathbf{I}^b_{x,y} - \mathbf{I}'^b_{x,y}\right|. \tag{1}$$

For each dot $\mathbf{c} \in \mathcal{C}$, we compute the total visual mismatch over its anti-aliased boundary pixels $\mathcal{S}$ (defined next) as:

$$\ell = \sum_{(x,y)\in\mathcal{S}} d(\mathbf{I}_{x,y}, \mathbf{I}'_{x,y}). \tag{2}$$

This focuses optimization on the boundary regions where anti-aliasing encodes fine-grained positional information.

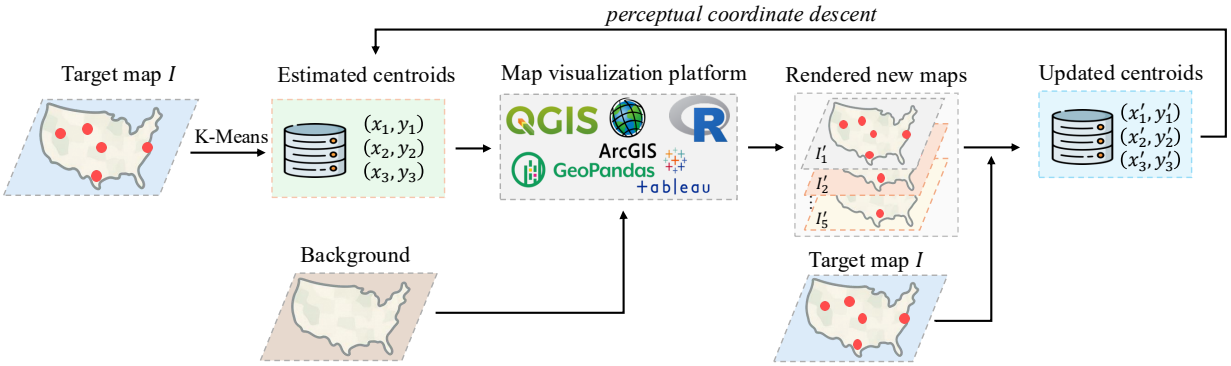


Figure 2: Illustration of the key processes in perceptual boundary alignment (PBA). The adversary first initializes dot centroids using K-Means clustering. Using map visualization tools and the map background, the adversary generates new maps based on the current centroid estimates, exploring five optimization directions (*i.e.,* stay, left, right, up, and down). The adversary then selects the new centroids that minimize the perceptual loss, which is computed by comparing the visual mismatch between generated maps and the target map. This process is repeated to improve centroid estimates.

Detecting Anti-Aliased Boundary Pixels. To compute the loss, we need to identify the set of anti-aliased boundary pixels $\mathcal{S}$ for each dot. These pixels, created by blending the dot and background colors, are the most sensitive to centroid changes. To identify these pixels, we perform a local boundary search around each estimated centroid. Specifically, starting from the centroid pixel, we use a breadth-first search (BFS) to find connected pixels whose color matches the dot color $\mathbf{z}$ (*i.e., inner pixels*). For each inner pixel, we examine its four-connected neighbors; neighbors whose color differs from $\mathbf{z}$ are collected as *anti-aliased boundary pixels*, since their color results from blending the dot and the background. These boundary pixels form the set $\mathcal{S}$ used to compute the perceptual loss in Equation (2).

While the above boundary detection works well for isolated dots, it becomes challenging for overlapping dots, where shared edges may appear as inner pixels of a merged region. To handle overlaps, we restrict the search to a region within the shape of the dot, which can be computed using its shape ϕ and size ρ. This ensures that $\mathcal{S}$ contains only the boundaries relevant to the current dot, avoiding contamination from neighboring ones. This procedure accurately identifies anti-aliased boundaries for both isolated and overlapping dots, providing a stable signal for exploiting anti-aliasing artifacts.

Perceptual Coordinate Descent (PCD). While we can compute the perceptual loss for the current estimate of each dot's centroid using boundary pixels, applying standard optimization methods such as stochastic gradient descent (SGD) [51], [52] to minimize this loss is infeasible. This is because the map rendering process is a black-box, making it hard to compute gradients. To address this, we propose Perceptual Coordinate Descent (PCD), a gradient-free algorithm inspired by zeroth-order optimization that is widely used in machine learning [53]–[55]. Instead of explicitly computing gradients, PCD probes nearby coordinates of centroids in the two-dimensional pixel space and selects the move that most reduces the perceptual loss, using the anti-aliased boundary pixels as a high-sensitivity signal. This process is repeated until no direction produces further improvement.

Framework Overview. The complete location recovery framework based on perceptual coordinate descent (PCD) is outlined in Algorithm 2 and Figure 2. The algorithm consists of two main phases. Phase 1 (lines 1-8) estimates the initial centroids of all dots. The algorithm first identifies all connected pixel clusters matching the dot color. For each cluster, it estimates the number of dots k it contains based on the cluster's size relative to a single dot's pixel area (μ). The K-Means clustering algorithm [56] is then applied to find k initial centroids. For an isolated dot, this simplifies to the baseline estimation in Algorithm 1. Phase 2 (lines 9-37) iteratively refines initial centroids using the proposed PCD. In each iteration, the algorithm considers five candidate directions (*i.e.,* stay, left, right, up, and down) on the two-dimensional pixel grid, with a step size η. It generates five candidate maps by shifting all current centroids $\mathcal{C}$ in these directions and rendering the corresponding maps (lines 14-17). For each dot, the perceptual loss is computed by evaluating the dot's anti-aliased boundary pixels in each direction (lines 20-33). The centroid that yields the smallest perceptual loss is selected for each dot. This process is repeated for T iterations to ensure convergence.

4.2. Geometric Refinement (GR)

We now consider the attack scenario where the adversary does not possess the background of the dot map. Without the background, the adversary cannot render comparison maps, so the optimization-based approach from the previous section is no longer applicable. To address this, we propose two complementary strategies to exploit anti-aliasing artifacts without the knowledge of the background: a weighted centroid estimation for isolated dots and a geometric curve fitting method for overlapping dots.

Isolated Dots: Refinement via Weighted Centroid. We begin with the same initial centroid estimate as in Phase 1

Algorithm 2 Perceptual Boundary Alignment (PBA). The algorithm applies the proposed perceptual coordinate descent (detailed in Section 4.1) to iteratively refine initial dot centroids by minimizing the perceptual loss computed on the dots' anti-aliased boundary pixels $\mathcal{S}$.

Require: Dot map $\mathbf{I}$, background $\mathbf{B}$, map rendering function $\mathcal{R}$, minimum number of pixels in a dot μ, dot color $\mathbf{z}$, dot geometry ϕ, dot size ρ, iterations T, step size η

```
# Phase 1: Initialize dots' centroids
P ← FindPixelClusters(I, z)
C ← ∅                                   ▷ initialize centroids
for each cluster P ∈ P do
    k = ⌈|P| / μ⌉                       ▷ estiamte # dots in this cluster
    C_P ← K-Means(P, k)
    C ← C ∪ C_P
end for
# Phase 2: Perceptual descent
# define search coordinates (stay, left, right, up, down)
D ← {(0,0), (η,0), (−η,0), (0,η), (0,−η)}
for T times do
    # Render new maps for each coordinate direction
    for each d ∈ D do
        C_d ← {c + d | c ∈ C}           ▷ shift all centroids by d
        I_d ← R(B, C_d, z, φ, ρ)
    end for
    C_new ← ∅                           ▷ initialize updated centroids
    # evaluate loss for each dot across rendered maps
    for each dot centroid c ∈ C do
        ℓ_min ← +∞
        c* ← c
        for each d ∈ D do
            c' ← c + d                  ▷ candidate centroid position
            I' ← I_d                    ▷ get rendered map for this shift
            S ← BoundarySearch(I, c', z, φ, ρ)
            ℓ' = Σ_{(x,y)∈S} d(I_{x,y}, I'_{x,y})   ▷ compute loss
            if ℓ' < ℓ_min then
                ℓ_min ← ℓ'
                c* ← c'
            end if
        end for
        C_new ← C_new ∪ {c*}
    end for
    C ← C_new                           ▷ update all centroids for next iteration
end for
return C
```

of Algorithm 2. Let $\mathcal{I}$ denote the set of inner pixels whose color matches the dot color $\mathbf{z}$, and let $\mathcal{S}$ represent the set of anti-aliased boundary pixels that can be detected using the same method used in perceptual boundary alignment. For each boundary pixel $(x, y) \in \mathcal{S}$, we approximate its local background color by averaging the colors of its neighbors:

$$\mathbf{b}_{x,y} = \frac{1}{|\mathcal{N}_{x,y}|} \sum_{(u,v)\in\mathcal{N}_{x,y}} \mathbf{I}_{u,v},$$

where $\mathcal{N}_{x,y}$ denotes the eight nearest neighboring pixels that are not part of the dot region (*i.e.,* neither inner nor boundary pixels). Under a linear blending assumption, the observed color of an anti-aliased pixel is a weighted mixture of the dot color $\mathbf{z}$ and the estimated background color $\mathbf{b}_{x,y}$. We estimate the dot's fractional contribution to each boundary pixel (x, y) as:

$$w_{x,y} = \frac{d(\mathbf{b}_{x,y}, \mathbf{I}_{x,y})}{\max(d(\mathbf{b}_{x,y},\ \mathbf{z}), d(\mathbf{b}_{x,y}, \mathbf{I}_{x,y}))},$$

where $d(\cdot, \cdot)$ is the color distance function defined in Equation (2). Intuitively, this weight (ranging from 0 to 1) estimates the dot's rendering influence on pixel (x, y) by calculating how far the pixel's color is from the background, normalized by the total possible color distance. For all interior pixels $(x, y) \in \mathcal{I}$, we set $w_{x,y} = 1$. The refined centroid is then computed as the blend-weighted center of mass over all contributing pixels:

$$(x_c, y_c) = \frac{\sum_{(x,y)\in\mathcal{I}\cup\mathcal{S}} w_{x,y} \times (x, y)}{\sum_{(x,y)\in\mathcal{I}\cup\mathcal{S}} w_{x,y}}.$$

This approach effectively shifts the centroid toward boundary pixels with stronger dot influence, dramatically improving location recovery accuracy without requiring background access or additional rendering.

Overlapping Dots: Refinement via Geometric Fitting. While the weighted-centroid method is highly effective for isolated dots, it fails in the presence of overlapping dots. This is because in overlapping regions, many boundary pixels are shared between neighboring dots, making it impossible to isolate a complete and unbiased set of boundary pixels for the target dot. Recall that most dot maps in literature consist of isolated dots, and accurately recovering their locations already presents a significant privacy risk. Despite this, we propose the following algorithm to handle overlapping dots. We adopt a geometric fitting approach that leverages the partial boundary segments that can be confidently attributed to each dot. Specifically, we first perform the same local boundary search as described in Algorithm 2 to extract the anti-aliased pixels belonging to the target dot. These pixels, although only a subset of the full boundary, form a reliable *segment* along the dot's perimeter. Using the known geometry of the dot, we can estimate the centroid based on this segment. For instance, if the dot is circular, the segment would correspond to an arc of the circle. Given that the dot's size (ρ) is known, we can apply a geometric fitting algorithm (*e.g.,* a circle-fitting algorithm [57]) to find the unique circle that best passes through this partial arc. The center of the fitted circle is then used as the refined centroid. This "fit-to-segment" approach can also be generalized to other common dot shapes using different geometric fitting algorithms [58].

The weighted centroid and geometric fitting algorithms provide a complementary location recovery refinement pipeline for the adversary without the map background. As shown in Section 5, these methods achieve strong recovery accuracy while maintaining high efficiency.

5. Evaluation

We conduct a comprehensive evaluation of AutoLocate across various attack settings to assess the privacy risks associated with different types of dot maps. Specifically, we aim to address the following research questions:

- **RQ1:** How effective is our proposed location recovery framework compared to existing approaches across different configurations of dot maps?
- **RQ2:** How do the different components of AutoLocate impact location recovery performance? How efficient is our approach?
- **RQ3:** What factors influence the location recovery accuracy of the dots?

5.1. Experimental Setup

Target Dot Map Construction. To the best of our knowledge, no public datasets or benchmarks exist for the dot map location recovery task. In addition, using dot maps from existing publications or articles would raise significant privacy and ethical concerns and lack a reliable ground truth for evaluating attack performance. To address these issues, we design the following pipeline to generate synthetic dot maps for our experiments:

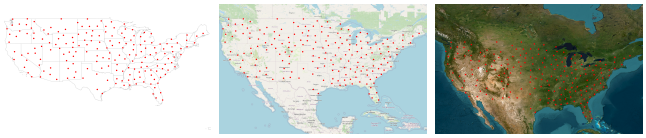
1) *Map and Dot Configurations.* We first configure different map settings (*e.g.,* geographical range, scale, and background) and dot settings (*e.g.,* color, geometry, and dot size) to simulate dot maps with different styles.
2) *Dot Generation.* For isolated dots, we randomly generate 188 high-precision latitude and longitude coordinates (with seven decimal places) by uniformly sampling locations within the geographical range of the map. For overlapping dots, we begin by sampling a location, then randomly perturb its coordinates to ensure the new location is closely connected to the original. We repeat this process to create dot clusters with 2, 3, or 4 dots, generating 25 clusters for each configuration.
3) *Map Composition and Export.* Using a popular map visualization platform (detailed below), we input the generated dot locations and map/dot configurations, compose the map, and export it in a common image format with varied resolutions. The resulting raster dot map is the target map from which we aim to recover the locations.

The benefits of using synthetic dot maps are twofold: (i) it provides a controllable setting to systematically study the privacy risks of dot maps by varying configurations (*e.g.,* visualization software, scale, dot types); and (ii) it avoids the ethical issues of using real-world sensitive data. We will release our dataset construction code, the generated target maps, and their corresponding ground-truth locations to facilitate future research on dot map privacy.

Map Scales and Background. We consider three map scales: national, state, and city level. We select a specific region for each: the United States (national), Ohio (state), and Austin, Texas (city), as shown in Figure 3. For each scale, we also test three different background types: a blank white canvas, a street map from OpenStreetMap [50], and a satellite map provided by Esri [59]. A demonstration of the generated national maps with different backgrounds is shown in Figure 4.

(a) Austin map (b) Ohio map (c) U.S. map

Figure 3: Dot maps at different geographic scales: (a) Austin, TX (city level); (b) Ohio (state level); and (c) the United States (national level).



(a) White canvas (b) Street map (c) Satellite map

Figure 4: Dot maps at the national scale (*i.e.,* United States) with different backgrounds: (a) a blank white canvas, street map, (b) a street map, and (c) a satellite map. We only show the isolated dots for visual clarity.

Map Visualization Platforms and Map Formats. We use widely used map visualization platforms to generate target dot maps. Specifically, we select GeoPandas [9], QGIS [7], and R [10] (with the maps package [60]). These platforms are chosen because they are free to use and popular across various academic disciplines. We use the latest stable version of each platform (GeoPandas 1.1.1, QGIS 3.44, and R 4.5.2) and utilize their default map composition settings. For the map output, we select three common image formats: PNG, JPEG, and TIFF.

Baselines. We compare our attacks against the following location recovery algorithms:

- *PixelMatch.* This method uses full-color pixel matching for centroid estimation, as introduced in Section 3, to estimate the locations of the dots.
- *PixelAvg.* This method combines the inner pixels (*i.e.,* pixels matching the dot color) and boundary pixels (*i.e.,* pixels adjacent to the inner pixels). The mean location of these pixels is used to estimate each dot's centroid.
- *Raster2Vec.* QGIS provides a built-in tool that converts raster objects in a map into vector geometries, which has been used in prior work [18] for location recovery. The dot's location is then determined by computing the centroid of the resulting vector geometry.

Previous studies on location recovery have only focused on isolated dots. Therefore, we only compare our method against these baselines on the recovery error for isolated dots. Furthermore, we exclude early approaches that rely on manual visual inspection [15], [16], as such methods are subjective and not reproducible.

Evaluation Metrics and Process. We use the recovery error to assess the performance of location recovery algorithms and the privacy risks of dot maps. The error can be evaluated in the following ways:

- *Absolute Geographical Error.* For each attack, we estimate the dot's centroid in pixel coordinates. We then apply the coordinate transformation to convert the pixel locations back to latitude and longitude. The difference between the estimated and ground truth locations is computed, which gives the latitude error, the longitude error, and the geographical recovery error ($L2$ distance) in meters.
- *Relative Pixel Error.* We calculate the geographical error (in meters) and normalize it by the real-world distance that a single pixel represents at that map's scale. This provides a relative error at the pixel level.

We compute and report the average error across all dots, analyzing isolated dots and overlapping dots separately.

Hyperparamter Settings. The proposed perceptual boundary alignment (PBA) method includes several hyperparameters. To demonstrate its robustness, we use consistent settings across all experiments. Specifically, we set the number of iterations to $T = 30$ and the step size to $\eta = 0.05$ pixels, with learning rate decay [61] set to 0.75. To estimate the average number of pixels per dot (*i.e.,* μ), we randomly select five isolated dots and compute their mean pixel coverage using a simple full-color match; this μ is then used to estimate the number of dots within each pixel cluster. For rendering new maps in PBA, we set the map render function $\mathcal{R}$ to the same map tool used to generate the target map. In Section 5.3 we show that performance remains robust when the map rendering tool used for attack differs from the one used to generate the target map.

Attack Setup. We use the Pillow library [62] to load target map images, process pixel data, and run our recovery algorithms. All evaluated map visualization tools provide command-line interfaces or APIs, enabling automated map generation for the perceptual boundary alignment procedure. We also leverage the coordinate-transformation functions provided by these platforms to convert between pixel and geographic coordinates. The default target maps are at the national scale (*i.e.,* the United States), exported with PNG format with 192 DPI. Dots are rendered as circles, in red color (*i.e.,* $\mathbf{z} = (255, 0, 0)$) with a size ρ of 2 mm (approximately covering 15 pixels in this setting). The entire recovery pipeline is fully automated and requires no human intervention or visual inspection.

5.2. Evaluation of AutoLocate (RQ1)

Performance Across Different Map Scales. We used a street map background and GeoPandas to generate target maps at three different scales (national, state, and city) to evaluate location recovery performance. As shown in Table 2, for maps at the national scale, all baseline attacks (*i.e.,* PixelMatch, PixelAvg, and Raster2Vec) achieve very similar recovery errors on the order of hundreds of meters. This indicates they are ineffective for accurately identifying locations from dot maps at a large geographical scale. Interestingly, the performance of PixelAvg and Raster2Vec is identical. Upon manual inspection, we confirmed that the vector geometry produced by the Raster2Vec tool captures the exact same set of pixels as the PixelAvg method (*i.e.,* both inner and boundary pixels). Consequently, their centroid calculations and resulting performance are the same. In contrast, our proposed methods, which exploit anti-aliasing artifacts for enhanced location recovery, achieve errors as low as approximately 1 meter when the background is available (*i.e.,* PBA) and around 10 meters when it is not (*i.e.,* GR). These results represent more than $200\times$ and $20\times$ improvements in location recovery performance over the strongest baselines. This finding demonstrates that precise locations can still be recovered even for dot maps covering large geographic areas. This performance trend is consistent across all three scales, with all methods achieving a similar relative pixel error.

Performance Across Different Map Backgrounds. We vary the map background (*i.e.,* white canvas, street map, and satellite imagery) to examine its impact on recovery accuracy. As shown in Table 3, the baseline methods achieve similar performance across all backgrounds, as they rely solely on dot color and do not account for anti-aliasing effects introduced by the interaction between dots and background textures during rendering. For perceptual boundary alignment (PBA), performance remains consistent across different backgrounds because the adversary has access to the background, allowing anti-aliasing patterns to be effectively modeled during perceptual coordinate descent. The performance of geometric refinement (GR) decreases as the background becomes more complex (from white canvas to satellite). This occurs because complex backgrounds make local background estimation less accurate, reducing the reliability of the linear weighting mechanism used for centroid estimation. Nevertheless, GR still significantly outperforms all baselines, even on a complex satellite background.

Performance Across Map Resolution. We also evaluate the impact of image resolution on recovery performance by rendering the maps at three different resolutions: 384, 192, and 96 DPI. The results are presented in Table 4. As expected, the performance of all methods degrades with lower resolution due to the reduced number of pixels available for the adversary to exploit when estimating dot centroids. Despite this, our proposed algorithms maintain strong performance even at a relatively low resolution. At 96 DPI, our PBA method still achieves a recovery error of approximately 5 meters. This level of accuracy is still effective for precise location recovery and continues to significantly outperform the baselines under the same conditions.

Performance Across Dot Properties. We vary the dot properties (*i.e.,* geometry and size) to examine their impact on recovery accuracy. As shown in Table 5, the performance of all methods decreases when more complex dot shapes are used (*e.g.,* changing from circles to pentagons). This degradation is more pronounced for the baselines and Geometric Refinement (GR), since the baselines are not designed to

TABLE 2: Average location recovery error for isolated dots on street-map backgrounds, compared across different map scales. A " ✓ " indicates that the method requires access to the map background, whereas a " ✗ " indicates that it does not.

Map Scale	Method	Background	Lat Error	Lon Error	Recovery Error (meter)	Relative Pixel Error
National (1:10M)	PixelMatch	✗	0.001494	0.001352	227.96	0.079605
	PixelAvg	✗	0.001622	0.001488	251.00	0.087651
	Raster2Vec	✗	0.001622	0.001488	251.00	0.087651
	PBA (ours)	✓	**0.000006**	**0.000010**	**1.18**	**0.000412**
	GR (ours)	✗	0.000071	0.000059	10.17	0.003581
State (1:1M)	PixelMatch	✗	0.000144	0.000162	24.19	0.080342
	PixelAvg	✗	0.000181	0.000187	28.73	0.095423
	Raster2Vec	✗	0.000181	0.000187	28.73	0.095423
	PBA (ours)	✓	**0.000001**	**0.000001**	**0.10**	**0.000345**
	GR (ours)	✗	0.000012	0.000009	1.65	0.005519
City (1:100K)	PixelMatch	✗	0.000020	0.000016	3.11	0.081686
	PixelAvg	✗	0.000024	0.000020	3.73	0.100429
	Raster2Vec	✗	0.000024	0.000020	3.73	0.100429
	PBA (ours)	✓	**0.000000**	**0.000000**	**0.02**	**0.000549**
	GR (ours)	✗	0.000001	0.000001	0.19	0.005222

TABLE 3: Comparison of the average location recovery error for isolated dots across different map backgrounds.

Background	Method	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
White canvas	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000007**	**0.000010**	**1.35**	**0.0005**
	GR	0.000034	0.000044	5.77	0.0020
Street	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000006**	**0.000009**	**1.20**	**0.0004**
	GR	0.000071	0.000059	10.17	0.0036
Satellite	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000008**	**0.000010**	**1.39**	**0.0005**
	GR	0.000152	0.000180	25.58	0.0089

TABLE 4: Comparison of the average location recovery error for isolated dots across different map resolutions.

Resolution	Method	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
384 DPI (4568 × 2848)	PixelMatch	0.000382	0.000375	58.15	0.0406
	PixelAvg	0.000481	0.000459	73.00	0.0510
	Raster2Vec	0.000481	0.000459	73.00	0.0510
	PBA	**0.000003**	**0.000003**	**0.50**	**0.0003**
	GR	0.000019	0.000024	3.21	0.0023
192 DPI (2284 × 1424)	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000006**	**0.000010**	**1.18**	**0.0004**
	GR	0.000071	0.000059	10.17	0.0036
96 DPI (1142 × 712)	PixelMatch	0.003225	0.003291	503.73	0.0775
	PixelAvg	0.003075	0.003329	494.64	0.0761
	Raster2Vec	0.003075	0.003329	494.64	0.0761
	PBA	**0.000034**	**0.000038**	**5.41**	**0.0009**
	GR	0.000544	0.000587	88.77	0.0137

handle complex centroid geometries, and GR relies on a simplified geometric model that does not account for irregular shapes. Additionally, as the dot size increases, recovery accuracy generally improves. This is especially true for the baselines, which rely on a mean for centroid estimation. In comparison, our perceptual boundary alignment (PBA) remains effective even when the dot is very small (*i.e.,* 1 mm in size, covering only approximately 4 pixels).

TABLE 5: Comparison of the average location recovery error (in meters) across different dot geometries and sizes.

Geometry	Method	Dot size=1(mm)	Dot size=2(mm)	Dot size=3(mm)
Circle	PixelMatch	564.25	227.96	114.20
	PixelAvg	291.26	251.00	134.03
	Raster2Vec	291.26	251.00	134.03
	PBA	**1.67**	**1.18**	**0.72**
	GR	15.10	10.17	8.11
Triangle	PixelMatch	550.81	506.22	432.19
	PixelAvg	613.12	521.27	450.36
	Raster2Vec	613.12	521.27	450.36
	PBA	**4.42**	**4.23**	**3.85**
	GR	118.08	54.71	37.62
Pentagon	PixelMatch	371.20	367.03	346.I64
	PixelAvg	430.41	400.76	388.07
	Raster2Vec	430.41	400.76	388.07
	PBA	**8.35**	**6.38**	**5.16**
	GR	56.13	28.57	22.96

Performance Across Map Visualization Platforms. We evaluated our attack's robustness by testing it on dot maps generated from three popular map visualization platforms: GeoPandas, QGIS, and R. As shown in Table 6, we observe no significant difference in recovery error across the three platforms. It demonstrates that although these platforms may use different internal (unknown) map rendering processes, our attacks remain robust and achieve high-precision location recovery regardless of the visualization software.

Performance Across Image Formats. Dot maps can be exported in various image formats for publication or dissemination. We evaluate the performance of our attacks on dot maps exported in three common formats: PNG, JPEG, and TIFF. As shown in Table 7, recovery performance is similar for PNG and TIFF formats. This is expected, as all methods operate on the image's RGB channels, which are preserved across these formats. However, we observed a drop in performance for JPEG, particularly for the GR and baseline methods. We suspect this performance degradation is due to the compression losses inherent in the JPEG format, which result in a less accurate representation of the map. Despite this, our methods still outperform the baselines by nearly

TABLE 6: Comparison of the average location recovery error for isolated dots across different map platforms.

Platform	Method	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
GeoPandas	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000006**	**0.000010**	**1.18**	**0.0004**
	GR	0.000071	0.000059	10.17	0.00355
QGIS	PixelMatch	0.001380	0.001496	227.74	0.0794
	PixelAvg	0.0017952	0.001656	273.30	0.0953
	Raster2Vec	0.0017952	0.001656	273.30	0.0953
	PBA	**0.000008**	**0.000006**	**1.09**	**0.0004**
	GR	0.000090	0.000067	12.76	0.0044
R	PixelMatch	0.000969	0.000915	149.41	0.5217
	PixelAvg	0.000064	0.000065	192.53	0.0672
	Raster2Vec	0.000064	0.000065	192.53	0.0672
	PBA	**0.000008**	**0.000009**	**1.49**	**0.0005**
	GR	0.000064	0.000065	10.18	0.0036

TABLE 7: Comparison of the average location recovery error for isolated dots across different map formats.

Map Format	Method	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
PNG	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000006**	**0.000010**	**1.18**	**0.0004**
	GR	0.000071	0.000059	10.17	0.0036
JPEG	PixelMatch	0.002201	0.2097	335.83	0.1172
	PixelAvg	0.001989	0.001803	303.84	0.1060
	PixelAvg	0.001989	0.001803	303.84	0.1060
	PBA	**0.000009**	**0.000014**	**1.71**	**0.0006**
	GR	0.000227	000234	35.51	0.0124
TIFF	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000006**	**0.000010**	**1.18**	**0.0004**
	GR	0.000071	0.000059	10.17	0.0036

ten times in the JPEG format. These results demonstrate that our proposed methods generalize well across common raster image formats and can achieve high-precision location recovery, regardless of the map format.

Performance on Overlapping Dots. We also evaluated the recovery performance for overlapping dots on a national-scale map with a street-map background. The results are presented in Table 8. As expected, recovery accuracy decreases as the number of overlapping dots increases, since overlapping regions blur dot boundaries and make it harder to leverage anti-aliasing artifacts for accurate centroid estimation. However, the performance of our PBA method only slightly decreases, maintaining a recovery error of less than 2 meters even with four overlapping dots. In comparison, GR experiences a more significant drop in accuracy, with the recovery error increasing from 10 meters to 55 meters as the number of overlapping dots rises from 1 to 4. Nevertheless, even when our methods (*i.e.,* PBA and GR) are tasked with recovering locations from four overlapping dots, their accuracy still exceeds that of the baselines operating on the much simpler task of recovering single, isolated dots. These results demonstrate the robustness and effectiveness of AutoLocate, even under challenging overlapping scenarios.

It is worth noting that although the performance of our attacks degrades on overlapping dots, this does not imply a low privacy risk for dot maps. This is because most published maps (as demonstrated in Section 2) are dominated by isolated dots, for which our attacks (both PBA and GR) can recover their locations with unprecedented precision.

TABLE 8: Comparison of the average location recovery error for isolated dots (size 1) and overlapping dots (sizes 2–4) at the national scale with a street map background.

# Overlaps	Method	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
1	PixelMatch	0.001494	0.001352	227.96	0.0796
	PixelAvg	0.001622	0.001488	251.00	0.0877
	Raster2Vec	0.001622	0.001488	251.00	0.0877
	PBA	**0.000006**	**0.000010**	**1.18**	**0.0004**
	GR	0.000071	0.000059	10.17	0.0036
2	PBA	**0.000006**	**0.00009**	**1.24**	**0.0004**
	GR	0.000276	0.0002133	39.18667	0.0134
3	PBA	**0.000007**	**0.000009**	**1.37**	**0.0004**
	GR	0.000301	0.000302	48.61	0.0165
4	PBA	**0.000007**	**0.00009**	**1.39**	**0.0004**
	GR	0.000344	0.000353	55.92	0.0190

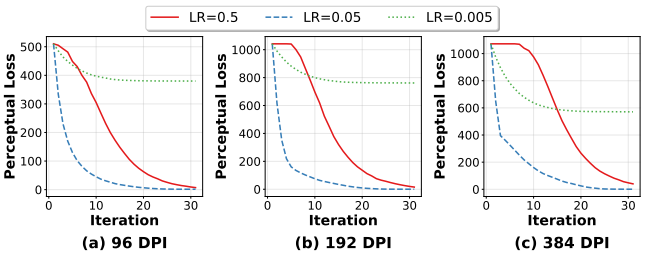


Figure 5: Impact of iterations and learning rate in PBA across different map resolutions.

5.3. Ablation Study (RQ2)

Impact of Optimization Hyperparameters. We analyze the impact of the optimization iterations T and step size η on the convergence and recovery performance of our perceptual boundary alignment (PBA) algorithm. Figure 5 shows the perceptual loss trajectories for national-scale maps at different resolutions using varied learning rates. We observe that a large learning rate (*i.e.,* 0.5) fails to converge and results in inaccurate estimations, while a small learning rate (*i.e.,* 0.005) leads to slow convergence. In general, we find that our default setting (*i.e.,* $\eta = 0.05$) reliably converges within 30 iterations and achieves the lowest perceptual loss. These hyperparameters consistently yield robust results across different map resolutions, demonstrating the stability of our algorithm.

Impact of Background in PBA. In perceptual coordinate descent, to compute perceptual loss, we generate a new map using the available background and the currently estimated dot centroids. However, the adversary may not have access to the exact background used in the target map. Here, we evaluate the impact of different background variations on recovery performance. Specifically, we select the street map from OpenStreetMap as the ground-truth background and create three scenarios for the background: (i) *Same*: The background is identical to the target map's background. (ii)

TABLE 9: Impact of map backgrounds for PBA.

Method	Background	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
PBA	Same	0.000006	0.000010	1.18	0.0004
	Perturb	0.000196	0.000205	30.49	0.1066
	Similar	0.000189	0.000150	27.25	0.0951
GR	-	0.00071	0.000059	10.17	0.0036

TABLE 10: Impact of map rendering platform for PBA.

Method	Rendering Platform	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
PBA	GeoPandas	0.000006	0.000010	1.18	0.000412
	QGIS	0.000017	0.000016	2.54	0.000887
	R	0.00015	0.00012	2.27	0.000792

Perturb: The background pixels are perturbed with Gaussian noise (*i.e.,* $\mathcal{N}(0, 10)$) on each RGB channel, visually maintaining the same appearance. (iii) *Similar*: A different street map from Esri [63] (instead of OpenStreetMap) for the same region, but from a different year.

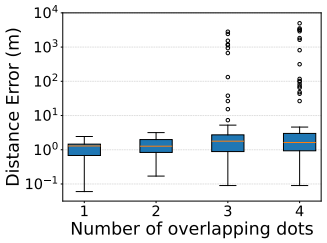
The results, compared with geometric refinement, are shown in Table 9. As shown, even small changes to the background (*i.e.,* perturbation) result in a drop in recovery accuracy. This is because PBA relies on analyzing subtle anti-aliasing artifacts at the dot boundaries, and any background change during rendering disrupts this process, leading to inaccurate loss calculations and ineffective optimization. In comparison, the performance of GR is not affected by the availability or changes of the background map. Therefore, we recommend using GR for location recovery when the background is unknown or cannot be precisely replicated.

Impact of Map Visualization Tools in PBA. In our previous experiments, we assumed that the adversary has access to the same map generation tool (*i.e.,* render function $\mathcal{R}$) as the target map. However, in practice, the adversary may not have this information. We test this scenario by fixing the target map as being generated by GeoPandas with a street map background. We then vary the tool the adversary uses to render the map during optimization. We test three different tools: GeoPandas, QGIS, and R. The results in Table 10 show that a misalignment of visualization tools between the target and the adversary has a minimal impact on recovery accuracy. This shows that our algorithm is robust even if the adversary does not know, or have access to, the specific tool used to generate the target map.

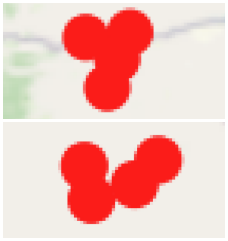
Efficiency Evaluation. We also evaluate the efficiency of our recovery algorithms and compare them with existing methods on a laptop with a Snapdragon X Elite CPU. As shown in Table 11, the baseline methods are very efficient, requiring less than one minute to estimate dot centroids on a map across different resolutions. Our geometric refinement approach is also efficient, though it is slightly slower than the baselines. In contrast, PBA requires more time (several minutes), with its runtime primarily dominated by the map composition process. Overall, the computing time (several minutes on a laptop) remains highly affordable. These results demonstrate that our attacks are practical and can be conducted by any adversary without requiring specialized computational resources.

TABLE 11: Comparison of the average running time (in minutes) for each attack.

Map Resolution	PixelMatch	PixelAvg	Raster2Vec	PBA	GR
1142 × 712	0.10	0.09	0.09	1.41	0.11
2284 × 1424	0.18	0.18	0.18	4.22	0.21
4568 × 2848	0.45	0.47	0.47	10.83	0.53



(a) Box plot illustrating the distribution of recovery errors across overlapping dots. Outliers are represented as circles.



(b) Top: Hard-to-attack overlapping dots. Bottom: Easy-to-attack dots.

Figure 6: (a) Analysis of recovery error distribution across varying numbers of overlapping dots. (b) Examples of easy-to-attack and hard-to-attack overlapping dots.

5.4. In-depth Analysis (RQ3)

While the previous experiments mainly examined privacy risks in dot maps by considering the average recovery error, privacy can also be viewed through a worst-case lens in many contexts [64], [65]. In this section, we delve deeper into the privacy risks of dot maps by analyzing recovery errors across individual dots, investigating why some dots are more vulnerable to attacks while others remain resistant.

Recovery Error Distribution. We first plot the distribution of recovery errors for all dots (including isolated and overlapping dots) on a national-scale map, as shown in Figure 6(a). From the plot, we observe that dots exhibit different recovery difficulty. While the median error is around 1 meter, some "easy-to-attack" instances (in the first quartile of the box plot) can be recovered with near-perfect accuracy. Conversely, "hard-to-attack" instances (above the third quartile) exhibit greater resilience against our attacks. Notably, overlapping dots, especially those with three or four overlaps, exhibit higher recovery errors compared to isolated dots. This suggests two key points: (i) the privacy risks associated with individual dots are not uniform across the map, and (ii) a large portion of dots in the map are vulnerable to our attacks because their recovery error is within a small range (*e.g.,* 1 meter). In the following analysis, we conduct case studies on both easy-to-attack and hard-to-attack dots to explore the underlying reasons for this disparity.

Why Some Dots are Hard to Attack? We select one representative hard-to-attack dot from the outliers in Figure 6(a). As seen in the top part of Figure 6(b), these dots are more tightly clustered together, which makes their individual anti-aliasing boundary pixels less distinct. In this case, it is difficult to estimate the visual mismatch between the generated map and the target map, as the perceptual loss

is calculated based on only a few valid boundary pixels. This is particularly true for the central dot, where its boundary pixels are surrounded by other dots, making it challenging to identify its precise location. As a result, reverse-engineering these dots becomes significantly more difficult, leading to higher recovery errors.

Why Some Dots are Easy to Attack? To understand why certain dots are more vulnerable, we select a representative example from the lower quartile of the recovery error distribution, focusing on dots with four overlaps. As shown in the bottom part of Figure 6(b), these dots are more separated, and their anti-aliasing boundary pixels are more complete with respect to their centroids. This makes it easier to compute the perceptual loss and measure the visual mismatch between the generated map and the ground truth. Additionally, the background of these dots is simpler, and the dot color contrasts sharply with the background. The lack of visual complexity in the background makes the anti-aliasing artifacts particularly prominent, providing a clear and high-fidelity signal for our attacks. This allows the adversary to estimate the centroids of these dots with extremely high precision.

6. Mitigation Strategies

Our findings suggest that even seemingly benign dot maps that cover a large geographic region can reveal concealed, high-precision location information through our attacks. Notably, in many maps, we were able to pinpoint a significant proportion of individuals' locations with accuracy within 1 meter. Thus, rigorous guidelines are needed to ensure the safe publication of dot maps. In this section, we discuss several potential mitigation strategies and propose a risk assessment tool for researchers to evaluate the privacy risks of dot maps.

Potential Mitigation Strategies. We consider the following mitigation strategies to defend against malicious location recovery attacks:

- *Publishing Maps Without Anti-Aliasing.* While anti-aliasing is enabled by default in all map visualization tools we are aware of, some tools (e.g., QGIS) allow the option to disable it. Since AutoLocate relies on anti-aliasing artifacts to reverse-engineer dot locations, disabling anti-aliasing can be an effective defense against our attacks.
- *Geo-masking.* A significant body of work [66]–[68] has developed various geo-masking techniques to mitigate privacy risks in location data. Here, we employ a simple approach [14], [69], which adds random noise to the raw location data by selecting a random direction within a fixed radius (50 or 100 meters).
- *Location Quantization.* Another straightforward mitigation involves quantizing the precision of the original high-precision location data (*e.g.,* reducing precision from seven decimal places to three or two). This fundamental change in the location's representation can protect the precise locations of individuals.

Note that these mitigation strategies were chosen because they are simple and widely known. While more sophisticated methods [14], [70], [71] exist, we focus on these methods to demonstrate the effectiveness of these mitigations against our attacks.

TABLE 12: Performance of mitigation strategies against the proposed attack (*i.e.,* PBA).

Mitigation	Lat Error	Lon Error	Dist. Error (m)	Rel. Px. Error
w/o Anti-alising	0.001082	0.001567	203.17	0.0710
Geo-masking (radius: 100 m)	0.000533	0.000792	100.02	0.0350
Geo-masking (radius: 50 m)	0.000286	0.000369	50.02	0.0175
Quantization (3 Decimals)	0.000239	0.000240	36.56	0.0128
Quantization (2 Decimals)	0.002579	0.002602	389.96	0.1362
-	0.000006	0.000010	1.18	0.00041

Mitigation Performance. We tested these mitigations on a national-scale map and applied our proposed attack (*i.e.,* PBA) to evaluate the recovery performance. As shown in Table 12, all mitigation strategies effectively degrade the recovery accuracy by a significant margin. Specifically, these mitigations increase the recovery error from around 1 meter to hundreds of meters, making the attack ineffective for identifying individuals within such a large area.

However, all of these mitigations come with some level of loss in map utility. For instance, forcing the map to render without anti-aliasing severely degrades the visualization quality, and we have not found published dot maps that currently adopt this method.

Privacy Risk Assessment Tool. Motivated by the above observations, we propose a privacy risk assessment tool to help researchers evaluate the privacy risks of their dot maps. We note that the success of location recovery depends on two factors: (1) the recovery accuracy of the attack and (2) the population density of the geographic region. In rural areas, even large errors may still reveal precise locations, whereas in dense urban areas, the same recovery error leads to higher uncertainty.

To address this, we develop a privacy assessment tool that adaptively adjusts the location quantization level based on local population density. Specifically, the tool takes the map configurations (*e.g.,* location data, map scale, and dot properties), our recovery attacks, and the population density of the region (which is publicly available, such as in [72], [73]) as input. It then outputs a recommended quantization level for the location data to ensure k-anonymity [74], where k is a user-defined parameter that specifies the desired level of privacy. This tool offers an intuitive and flexible way for researchers to balance privacy and map usability, while ensuring compliance with privacy regulations such as GDPR [75], CCPA [76], and HIPAA [77]. We will release the tool for free use upon publication.

7. Related Work

Dot maps are increasingly popular tools for visualizing the spatial distribution of individuals and events [3]–[5]. In academic publications and articles, dot maps are most commonly shared as raster images (*e.g.,* PNG, JPEG, and TIFF), reflecting the conventions of print media and the convenience of distributing fixed image formats [78]–[80].

Privacy Risks with Dot Maps. Dot maps are frequently used to display sensitive personal data, such as patient locations and crime incident locations. For example, Armstrong [81] highlighted that in epidemiological and criminal investigations, it is common for dot maps to have a one-to-one correspondence between each dot and a specific case. A significant body of research [12]–[14], [16], [18] demonstrates that these dots can be reverse-engineered to re-identify precise locations, posing serious privacy risks. For instance, Curtis et al. [13] successfully identified the locations of Hurricane Katrina fatalities published in The Baton Rouge Advocate, showing that residential identification can be achieved with minimal geographic references. Similarly, Brownstein et al. [18] found that over 26% of locations from presentation-quality maps and over 79% from publication maps could be accurately identified. Kounadi et al. [12] identified 41 articles between 2005 and 2012 that disclosed over 68,000 home addresses, emphasizing the scope of identification risks. These studies raise ethical and security concerns, especially for individuals with stigmatized conditions (*e.g.,* mental illness), as they could be targeted. In addition, burglars could identify vacant homes [12], [13], and insurers might infer health risks from this data.

Most existing work focuses on maps that cover a small geographic area, with little research addressing the privacy risks of dot maps that span broader regions (e.g., national-level maps). Such maps have been created for regions including Germany and the Netherlands [4], Cameroon [82], Thailand [83], and the United States [3], as listed in [3]. While dot maps covering larger regions are not uncommon, there is still a lack of systematic analysis regarding their privacy risks. Our work fills this gap by systematically assessing the privacy risks associated with dot maps at different geographic extents. Furthermore, existing geo-location privacy studies have not explored the use of anti-aliasing for location recovery. This is a key focus of our work, where we investigate how these map rendering techniques can be leveraged to recover high-precision location information.

Privacy Protection Strategies for Locations and Maps. Many studies propose geo-masking strategies to mitigate privacy risks in location-based data. One early approach is dot aggregation, where dot locations are aggregated at either the midpoint of the street segment or at the nearest street intersection [12], [66]. Another common technique is random perturbation, which introduces random noise to location coordinates. Various perturbation methods have been studied, including random direction and fixed radius [14], [69], random perturbation within a circle [84], [85], Gaussian displacement [85], [86], donut masking [67], [87], and bimodal Gaussian displacement [68]. Several studies extend quantitative privacy notions, such as k-anonymity [74] and differential privacy [64], to geo-location data, and develop location-preserving techniques [70], [71], [88], [89].

Deploying these defenses for dot map publications requires understanding the trade-off between privacy implications and map usability. One needs to choose an appropriate defense level to satisfy privacy requirements while achieving good visualization readability. Thus, we introduce a risk assessment framework. By integrating our attacks with a population density map, researchers can select an appropriate coordinate quantization precision tailored to their specific privacy and utility requirements, offering a flexible trade-off between privacy and map usability.

8. Conclusion

In this paper, we systematically study the privacy risks associated with dot maps by proposing an automated location recovery framework named AutoLocate. AutoLocate includes two location recovery algorithms that exploit anti-aliasing artifacts of rasterized dot maps for precise location estimation, depending on whether the map background is available to the adversary. Extensive experiments across different map visualization tools, map scales, and dot properties demonstrate the effectiveness and robustness of the proposed methods. AutoLocate achieves a recovery error of around 1 meter for national-scale maps, which is over 200 times more accurate than previous methods. Given the severity of the threat, we also explore several mitigation strategies and introduce a privacy assessment tool to help researchers assess and mitigate the privacy risks of their dot maps. Our work has several limitations. First, our experiments were conducted on synthetic and controlled dot maps, so the attack performance on real-world dot maps remains unknown. Second, our attack is only applicable to dot maps, and the privacy risks of other map types (*e.g.,* trace maps) remain unexplored. Nevertheless, our work discovers a new attack vector for recovering highly precise location information from dot maps and offers new directions for analyzing the privacy risks of dot maps.

Ethics Considerations

Our research investigates the privacy risks associated with dot maps, specifically focusing on high-precision location recovery from rasterized maps. Since dot maps are widely used to visualize sensitive data (e.g., patient home addresses and crime locations), we recognize our responsibility to carefully assess the ethical implications of our findings. We have undertaken this assessment using the framework outlined in the Menlo Report, while adhering to the ethical guidelines set forth by IEEE S&P 2026.

Stakeholder-Based Analysis. This research involves several key stakeholders, each impacted by our findings in different ways:

- *Researchers and Map Creators.* Our primary audience consists of researchers and creators of dot maps. We provide these practitioners with a deeper understanding of the privacy risks in dot maps, along with a concrete tool for assessing the risks of their own maps. Additionally, we propose and validate mitigation strategies to address these risks.
- *Data Subjects.* The data subjects in this context are the individuals whose sensitive location data is visualized on

dot maps. In this paper, our experiments were conducted using synthetic datasets, and no specific individuals or proprietary dot maps were targeted. Furthermore, we believe it is important to raise awareness about these underlying privacy risks and prevent potential privacy threats to individuals in the future.

- *Map Software Developers.* The developers of map visualization platforms (*e.g.,* QGIS, GeoPandas, and R) are also stakeholders in this research, as our attack exploits a default rendering feature (*i.e.,* anti-aliasing) present in these platforms. By publishing this work, we aim to provide developers with insights to incorporate techniques that can mitigate such privacy risks.
- *Adversaries.* Our methods could be maliciously used by adversaries to identify individuals or specific locations from dot maps. However, it is important to note that these risks already existed prior to our research. We believe that by raising awareness of these risks, we can help mitigate broader privacy concerns. Additionally, we discuss effective mitigation strategies to minimize the likelihood of malicious use of this research.

Ethical Justification and Disclosure of Vulnerabilities. Dot maps are commonly used in sensitive domains such as healthcare, urban planning, and epidemiology, where the potential risks of exposing individuals' locations are significant. Given that these risks are not always well understood, we believe it is crucial to disclose the vulnerabilities associated with publishing dot maps. While we recognize that malicious actors could exploit our findings, we believe that proactively sharing this knowledge enables the research community to address these privacy risks before they are exploited in real-world scenarios. We also emphasize the critical role of responsible research in cybersecurity. By presenting our findings, we encourage the community to use our results to develop stronger privacy protections, rather than to exploit these vulnerabilities for malicious purposes.

Artifacts

Our artifact consists of a code repository that includes the following components:

- Source code for our proposed attack framework, which encompasses both perceptual boundary alignment and geometric refinement algorithms, as well as the baselines required to replicate the experiments in this paper.
- Synthetic dot maps with ground truth dot locations, to facilitate benchmarking and future research on map privacy.
- A proposed privacy assessment tool designed to help researchers evaluate and assess the privacy risks of dot maps, and recommend the appropriate location quantization level to meet specific privacy requirements.

A detailed `README.md` file within the repository provides step-by-step instructions for setting up the environment and running the experiments. The anonymized repository is available for review at: https://anonymous.4open.science/r/AutoLocate. To promote transparency and encourage further research on dot map privacy, we will release the code used in our study upon publication.

References

[1] A. Chandran and P. Roy, "Applications of geographical information system and spatial analysis in Indian health research: a systematic review," *BMC Health Services Research*, vol. 24, p. 1448, 2024.

[2] L. P. Clark, D. Zilber, C. Schmitt *et al.*, "A review of geospatial exposure models and approaches for health data integration," *Journal of Exposure Science & Environmental Epidemiology*, vol. 35, pp. 131–148, 2025.

[3] C. Smith, S. Le Comber, H. Fry, M. Bull, S. Leach, and A. Hayward, "Spatial methods for infectious disease outbreak investigations: Systematic literature review," *Eurosurveillance*, vol. 20, 2015.

[4] L. Soetens, S. Hahné, and J. Wallinga, "Dot map cartograms for detection of infectious disease outbreaks: an application to Q fever, the Netherlands and pertussis, Germany," *Eurosurveillance*, vol. 22, no. 26, p. 30562, 2017.

[5] B. F. Martinez, J. L. Annest, E. M. Kilbourne, M. L. Kirk, K. Lui, and S. M. Smith, "Geographic Distribution of Heat-Related Deaths Among Elderly Persons: Use of County-Level Dot Maps for Injury Surveillance and Epidemiologic Research," *JAMA*, vol. 262, no. 16, pp. 2246–2250, 1989.

[6] Esri, "ArcGIS," https://www.arcgis.com.

[7] "QGIS," https://qgis.org/.

[8] "Tableau: Business Intelligence and Analytics Software," https://www.arcgis.com.

[9] "GeoPandas," https://geopandas.org/.

[10] "The R Project for Statistical Computing," https://www.r-project.org/.

[11] A. Murad and B. F. Khashoggi, "Using GIS for disease mapping and clustering in Jeddah, Saudi Arabia," *ISPRS International Journal of Geo-Information*, vol. 9, no. 5, p. 328, 2020.

[12] O. Kounadi and M. Leitner, "Why Does Geoprivacy Matter? The Scientific Publication of Confidential Data Presented on Maps," *Journal of Empirical Research on Human Research Ethics*, vol. 9, no. 4, pp. 34–45, 2014.

[13] A. J. Curtis, J. W. Mills, and M. Leitner, "Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina," *International Journal of Health Geographics*, vol. 5, no. 1, p. 44, 2006.

[14] P. A. Zandbergen, "Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data," *Advances in Medicine*, vol. 2014, no. 1, p. 567049, 2014.

[15] J. S. Brownstein, C. A. Cassa, I. S. Kohane, and K. D. Mandl, "Reverse geocoding: concerns about patient confidentiality in the display of geospatial health data," *AMIA Annual Symposium Proceedings*, vol. 2005, p. 905, 2005.

[16] M. Leitner, J. W. Mills, and A. Curtis, "Can Novices to Geospatial Technology Compromise Spatial Confidentiality?" *KN - Journal of Cartography and Geographic Information*, vol. 57, no. 2, pp. 78–84, 2007.

[17] M. Leitner and A. Curtis, "A first step towards a framework for presenting the location of confidential point data on maps—results of an empirical perceptual study," *International Journal of Geographical Information Science*, vol. 20, no. 7, pp. 813–822, 2006.

[18] J. S. Brownstein, C. A. Cassa, I. S. Kohane, and K. D. Mandl, "An Unsupervised Classification Method for Inferring Original Case Locations from Low-resolution Disease Maps," *International Journal of Health Geographics*, vol. 5, no. 1, p. 56, 2006.

[19] A. Curtis, J. Mills, and M. Leitner, "Keeping an eye on privacy issues with geospatial data," *Nature*, vol. 441, no. 7090, p. 150, 2006.

[20] W. J. Leler, "Human vision, anti-aliasing, and the cheap 4000 line display," *ACM Siggraph Computer Graphics*, vol. 14, no. 3, pp. 308–313, 1980.

[21] H. Freeman, "Computer processing of line-drawing images," *ACM Computing Surveys (CSUR)*, vol. 6, no. 1, pp. 57–97, 1974.

[22] Wikipedia, "1854 Broad Street cholera outbreak," https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak, 2025.

[23] N. Buamithup, K. Intawong, and V. Punyapornwithaya, "Geographical Distribution, Spatial Directional Trends, and Spatio-Temporal Clusters of the First Rapid and Widespread Lumpy Skin Disease Outbreaks in Thailand," *Transboundary and Emerging Diseases*, vol. 2025, no. 1, p. 4900775, 2025.

[24] H.-H. Lin, S. Shin, J. Blaya, Z. Zhang, P. Cegielski, C. Contreras, L. Asencios, C. Bonilla, J. Bayona, C. Paciorek, and T. Cohen, "Assessing spatiotemporal patterns of multidrug-resistant and drug-sensitive tuberculosis in a South American setting," *Epidemiology and infection*, vol. 139, pp. 1784–93, 12 2010.

[25] J. Pearson, C. Jacobson, N. Ugochukwu, E. Asare, K. Kan, N. Pace, J. Han, N. Wan, R. Schonberger, and M. Andreae, "Geospatial analysis of patients' social determinants of health for health systems science and disparity research," *International anesthesiology clinics*, vol. 61, no. 1, pp. 49–62, 2023.

[26] N. Mimnagh, "Revisiting John Snow's Cholera Map: A Data Visualisation Case Study for Statistical Education," *arXiv preprint arXiv:2504.13970*, 2025.

[27] H. A. Kebede, M. M. Assen, and M. A. Sharew, "Crime Hotspot Analysis and Mapping Using Geospatial Technology in Dessie City, Ethiopia," *arXiv preprint arXiv:2501.00036*, 2024.

[28] J. Eck, S. Chainey, J. Cameron, and R. Wilson, "Mapping crime: Understanding hotspots," 2005.

[29] B. Wang, Q. Liu, J. Liu, and Y. Jiang, "Identifying key factors associated with commuting burden and modelling their non-linear relationships: the case study of Shenzhen, China," *Transportation Planning and Technology*, pp. 1–25, 2025.

[30] X. Lu, H. Yan, W. Li, X. Li, and F. Wu, "An algorithm based on the weighted network Voronoi Diagram for point cluster simplification," *ISPRS International Journal of Geo-Information*, vol. 8, no. 3, p. 105, 2019.

[31] M. Viljanen, L. Tostrams, N. Schoffelen, J. van de Kassteele, L. Marshall, M. Moens, W. Beukema, and W. Wamelink, "A joint model for the estimation of species distributions and environmental characteristics from point-referenced data," *Plos one*, vol. 19, no. 6, p. e0304942, 2024.

[32] J. Abellán, M. Martínez-Beneito, O. Zurriaga, G. Jorques, J. Ferrándiz, and A. López-Quílez, "Point processes as a tool for analyzing possible sources of contamination," *Gaceta Sanitaria*, vol. 16, no. 5, pp. 445–449, 2002.

[33] M. Papeş and P. Gaubert, "Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents," *Diversity and distributions*, vol. 13, no. 6, pp. 890–902, 2007.

[34] C. S. Montalvo-Mancheno, J. C. Buettel, S. Ondei, and B. W. Brook, "A Reproducible, Data-Driven Approach to Mapping Species Distributions Using Presence-Only Data and Biogeographic Templates," *Ecology and Evolution*, vol. 15, no. 10, p. e72285, 2025.

[35] A. Dmowska and T. F. Stepinski, "Racial dot maps based on dasymetrically modeled gridded population data," *Social Sciences*, vol. 8, no. 5, p. 157, 2019.

[36] A. Dmowska and T. Stepinski, "Mapping racial diversity using grid-based racial dot maps and racial diversity maps," *Population Association of America*, 2019.

[37] K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, and E. Shook, "Mapping the global Twitter heartbeat: The geography of Twitter," *First Monday*, 2013.

[38] F. Agostinelli, M. Luflade, and P. Martellini, "On the spatial determinants of educational access," National Bureau of Economic Research, Tech. Rep., 2024.

[39] X. Yuan, "The Application of Geographic Information System (GIS) in Academic Success Center (ASC) of a Medium-Sized Liberal Art University." *Educational Research: Theory and Practice*, vol. 31, no. 3, pp. 94–100, 2020.

[40] M. A. Jochim, "Dots on the map: Issues in the archaeological analysis of site locations," *Journal of Archaeological Method and Theory*, vol. 30, no. 3, pp. 876–894, 2023.

[41] C. Keller, "Distribution of Badorf and Walberberg Ware in the British Isles during the 8th and 9th centuries," https://zenodo.org/records/10013458, 2023.

[42] Esri, "Understanding Coordinate Management in the Geodatabase," https://support.esri.com/en-us/technical-paper/understanding-coordinate-management-in-the-geodatabase-1301, 2007.

[43] Wikipedia, "Shapefile," https://en.wikipedia.org/wiki/Shapefile, 2025.

[44] V. Kesten, "Evaluating Different Spatial Anti Aliasing Techniques," 2017.

[45] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, "Modeling the interaction of light between diffuse surfaces," *ACM SIGGRAPH computer graphics*, vol. 18, no. 3, pp. 213–222, 1984.

[46] F. C. Crow, "The aliasing problem in computer-generated shaded images," *Communications of the ACM*, vol. 20, no. 11, pp. 799–805, 1977.

[47] "Anti-aliasing techniques comparison," https://www.sapphirenation.net/anti-aliasing-comparison-performance-quality, 2016.

[48] "A quick overview of MSAA," https://mynameismjp.wordpress.com/2012/10/24/msaa-overview/, 2012.

[49] T. Lottes, "A quick overview of MSAA," https://developer.download.nvidia.com/assets/gamedev/files/sdk/11/FXAA_WhitePaper.pdf, 2009.

[50] O. contributors, "OpenStreetMap," https://www.openstreetmap.org, 2024.

[51] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[52] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[53] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE transactions on automatic control*, vol. 37, no. 3, pp. 332–341, 2002.

[54] H. J. Kushner and D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*, 2012, vol. 26.

[55] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, p. 265–272.

[56] J. B. McQueen, "Some methods of classification and analysis of multivariate observations," in *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, 1967, pp. 281–297.

[57] I. D. Coope, "Circle fitting by linear and nonlinear least squares," *Journal of Optimization theory and applications*, vol. 76, no. 2, pp. 381–388, 1993.

[58] S. Arlinghaus, *Practical handbook of curve fitting*, 2023.

[59] Esri, "World Imagery," https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9, 2016.

[60] "maps: Draw Geographical Maps," https://cran.r-project.org/web/packages/maps/index.html.

[61] A. Krogh and J. Hertz, "A simple weight decay can improve generalization," *Advances in neural information processing systems*, vol. 4, 1991.

[62] "Pillow," https://pillow.readthedocs.io/.

[63] Esri, "Esri World Topographic Map," https://www.arcgis.com/home/item.html?id=6e850093c837475e8c23d905ac43b7d0, 2022.

[64] C. Dwork, "Differential Privacy," in *Automata, Languages and Programming*, 2006, pp. 1–12.

[65] N. Li, W. Qardaji, D. Su, Y. Wu, and W. Yang, "Membership privacy: A unifying framework for privacy definitions," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 889–900.

[66] M. Leitner and A. Curtis, "Cartographic guidelines for geographically masking the locations of confidential point data," *Cartographic Perspectives*, no. 49, pp. 22–39, 2004.

[67] D. Stinchcomb, "Procedures for geomasking to protect patient confidentiality," in *ESRI international health GIS conference*, 2004, pp. 17–20.

[68] C. A. Cassa, S. J. Grannis, J. M. Overhage, and K. D. Mandl, "A context-sensitive approach to anonymizing spatial surveillance data: impact on outbreak detection," *Journal of the American Medical Informatics Association*, vol. 13, no. 2, pp. 160–165, 2006.

[69] M.-P. Kwan, I. Casas, and B. Schmitz, "Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks?" *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 39, no. 2, pp. 15–28, 2004.

[70] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *2011 IEEE symposium on security and privacy*. IEEE, 2011, pp. 247–262.

[71] V. Bindschaedler and R. Shokri, "Synthesizing plausible privacy-preserving location traces," in *2016 IEEE symposium on security and privacy (SP)*. IEEE, 2016, pp. 546–563.

[72] "World Population Density," https://luminocity3d.org/WorldPopDen/.

[73] "GPWv411: Population Density," https://developers.google.com/earth-engine/datasets/catalog/CIESIN_GPWv411_GPW_Population_Density.

[74] L. Sweeney, "k-anonymity: A model for protecting privacy," *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 10, no. 05, pp. 557–570, 2002.

[75] P. Regulation, "Regulation (EU) 2016/679 of the European Parliament and of the Council," *Regulation (EU)*, vol. 679, 2016.

[76] S. of California Legislature., "California Consumer Privacy Act of 2018," *Public law*, 2018.

[77] A. Act, "Health insurance portability and accountability act of 1996," *Public law*, vol. 104, p. 191, 1996.

[78] R. K. Matiashuk, I. V. Goncharenko, I. V. Tkachenko, Y. S. Prokopuk, and K. Y. Shchur, "Taxonomic composition and spatial structure of the elements of the Feofaniya park," *Ecology and Noospherology*, vol. 26, no. 3-4, pp. 21–29, 2015.

[79] N. Dziuba and S. Szombara, "Supporting the editing of dot maps using the spectral clustering algorithm," *Polish Cartographical Review*, vol. 57, no. 1, pp. 58–74, 2025.

[80] N. Koktavá and J. Horák, "Options for micro-mobility data visualization," *European Journal of Geography*, vol. 14, no. 4, pp. 46–52, 2023.

[81] M. Armstrong, "Geographic information technologies and their potentially erosive effects on personal privacy," *Studies in the Social Sciences*, vol. 27, 01 2002.

[82] M. A. Tewara, P. N. Mbah-Fongkimeh, A. Dayimu *et al.*, "Small-area spatial statistical analysis of malaria clusters and hotspots in Cameroon; 2000–2015," *BMC Infectious Diseases*, vol. 18, no. 1, p. 636, 2018.

[83] K. Y. Maulana, K. Na-Lampang, O. Arjkumpa, N. Buamithup, K. Intawong, and V. Punyapornwithaya, "Geographical Distribution, Spatial Directional Trends, and Spatio-Temporal Clusters of the First Rapid and Widespread Lumpy Skin Disease Outbreaks in Thailand," *Transboundary and Emerging Diseases*, vol. 2025, no. 1, p. 4900775, 2025.

[84] M. P. Armstrong, G. Rushton, and D. L. Zimmerman, "Geographically masking health data to preserve confidentiality," *Statistics in medicine*, vol. 18, no. 5, pp. 497–525, 1999.

[85] D. L. Zimmerman and C. Pavlik, "Quantifying the effects of mask metadata disclosure and multiple releases on the confidentiality of geographically masked health data," *Geographical analysis*, vol. 40, no. 1, pp. 52–76, 2008.

[86] C. A. Cassa, S. C. Wieland, and K. D. Mandl, "Re-identification of home addresses from spatial locations anonymized by Gaussian skew," *International journal of health geographics*, vol. 7, no. 1, p. 45, 2008.

[87] Y. Lu, C. Yorke, and F. B. Zhan, "Considering risk locations when defining perturbation zones for geomasking," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 47, no. 3, pp. 168–178, 2012.

[88] K. El Emam, F. K. Dankar, R. Issa, E. Jonker, D. Amyot, E. Cogo, J.-P. Corriveau, M. Walker, S. Chowdhury, R. Vaillancourt *et al.*, "A globally optimal k-anonymity method for the de-identification of health data," *Journal of the American Medical Informatics Association*, vol. 16, no. 5, pp. 670–682, 2009.

[89] S. C. Wieland, C. A. Cassa, K. D. Mandl, and B. Berger, "Revealing the spatial distribution of a disease while preserving privacy," *Proceedings of the National Academy of Sciences*, vol. 105, no. 46, pp. 17 608–17 613, 2008.

Appendix A.
Literature Survey of Dot Map Usage

To demonstrate the prevalence of dot map usage across different research fields, we conducted a literature review using Google Scholar and Semantic Scholar academic search engines. Our search methodology combined general terms for the visualization technique (*e.g.,* "dot map", "point map", "spatial distribution", "geocoded data") with keywords specific to each research field and its corresponding sensitive data. The queries for each domain were structured as follows:

- For **Public Health**, we used ("public health" OR "epidemiology") AND "dot map" AND ("patient location" OR "case distribution").
- For **Criminology**, we used "crime mapping" AND "point map" AND "incident location".
- For **Urban Planning**, we used "urban planning" AND "spatial distribution" AND ("home-work location" OR "commute").
- For **Ecology**, we used "ecology" AND "point map" AND ("endangered species" OR "presence-only data").
- For **Social Science**, we used ("social science" OR "demography") AND "dot density map" AND ("household demographics" OR "racial dot map").
- For **Education**, we used "education" AND "spatial analysis" AND "student residence".
- For **Archaeology**, we used "archaeology" AND "point map" AND "artifact find spot".

For the publications found in the search engines, we manually checked each one to ensure that the paper included dot maps. If a publication was not a direct match, we expanded our search by examining its reference list to identify related studies. The resulting selected publications are shown in Table 1. Note that this is an exemplary list

intended to showcase the prevalence of dot map usage, not a comprehensive or systematic review. We refer the reader to [3] for a more complete survey of spatial visualization usage in academic research.