Stop Misusing t-SNE and UMAP for Visual Analytics

Hyeon Jeon
hj@hcil.snu.ac.kr
Seoul National University
Seoul, Korea

Jeongin Park
jipark@hcil.snu.ac.kr
Seoul National University
Seoul, Korea

Sungbok Shin
sungbok.shin@inria.fr
Inria-Saclay
Université Paris-Saclay
Scalay, France

Jinwook Seo
jseo@snu.ac.kr
Seoul National University
Seoul, Korea

Abstract

Misuses of *t*-SNE and UMAP in visual analytics have become increasingly common. For example, although *t*-SNE and UMAP projections often do not faithfully reflect the original distances between clusters, practitioners frequently use them to investigate inter-cluster relationships. We investigate why this misuse occurs, and discuss methods to prevent it. To that end, we first review 136 papers to verify the prevalence of the misuse. We then interview researchers who have used dimensionality reduction (DR) to understand why such misuse occurs. Finally, we interview DR experts to examine why previous efforts failed to address the misuse. We find that the misuse of *t*-SNE and UMAP stems primarily from limited DR literacy among practitioners, and that existing attempts to address this issue have been ineffective. Based on these insights, we discuss potential paths forward, including the controversial but pragmatic option of automating the selection of optimal DR projections to prevent misleading analyses.

CCS Concepts

• **Human-centered computing** → *Visual analytics*; • **Mathematics of computing** → **Dimensionality reduction**.

Keywords

t-SNE, UMAP, Dimensionality Reduction, Visual Analytics, Literature Review, Interview Study

ACM Reference Format:
Hyeon Jeon, Jeongin Park, Sungbok Shin, and Jinwook Seo. 2025. Stop Misusing t-SNE and UMAP for Visual Analytics. In . ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

We look closely into a widely known phenomenon threatening the reliability of visual analytics: *the misuse of t-SNE and UMAP.* When practitioners refer to dimensionality reduction (DR) for visually analyzing high-dimensional data, *t*-SNE and UMAP are the most frequently applied techniques [11, 19, 34, 47, 75]. However, these two techniques are also commonly misused in practice [8, 9, 16, 72]. For example, although *t*-SNE and UMAP do not accurately represent global structures like distances between points [19, 33, 56, 72], they are often used to investigate the dissimilarity between data points or clusters [8, 9, 72]. Such misuse may introduce errors in visual analytics, which can further propagate through interconnected visualizations in the systems and thus compromise their reliability.

We systematically investigate this misuse to understand how to address it. First, we verify the existence of the misuse by reviewing 136 visual analytics papers that utilize DR. Then, we conduct interviews with 12 researchers who frequently use DR for visual analytics purposes, including data visualization, machine learning, and bioinformatics, to detail the underlying causes of the misuse. As a final step, we interview eight DR researchers to understand why previous attempts have hardly addressed the misuse.

Our findings indicate that the misuse mainly occurs because practitioners have limited literacy of DR. For instance, several participants in our first interview study perceive *t*-SNE and UMAP to be "immune to criticism," implying that both authors and reviewers lack sufficient knowledge on how to use DR appropriately. Researchers across various domains have also sought to address this problem by producing many papers that warn the weakness of *t*-SNE and UMAP and that emphasize how to use DR properly (Sect. 2.2). However, these approaches have proven ineffective, failing to motivate practitioners to engage with these materials for enhancing DR literacy.

As a possible solution, we propose—although with some reluctance, to delegate the selection of DR projections to machines. We suggest automating the selection of optimal DR techniques and hyperparameters for given analytical tasks and contexts. These technical solutions enable practitioners to use DR effectively even without sufficient DR expertise, but may undermine user agency and discourage them from cultivating their DR literacy. We discuss how to preserve agency while simultaneously preventing misuse of DR techniques.

In summary, our research provides three key contributions:

- We present a **literature review** and **two interview studies** that investigate the misuse of *t*-SNE and UMAP.
- We verify the **existence of misuses** and understand why the existing approaches are ineffective in addressing the misuse.
- We suggest delegating the selection of optimal DR projections to automated systems as a viable solution and discuss the pros and cons of this approach.

We hope this research will spark discussions and encourage more appropriate use of not only DR but also other machine learning techniques, ultimately enhancing the reliability of visual analytics.

2 Backgrounds and Related Work

We first present how the two DR techniques, *t*-SNE and UMAP, impact visual analytics research. We then detail previous efforts in visualization, machine learning, and bioinformatics fields to address this problem.

Conference, Washington, DC, USA

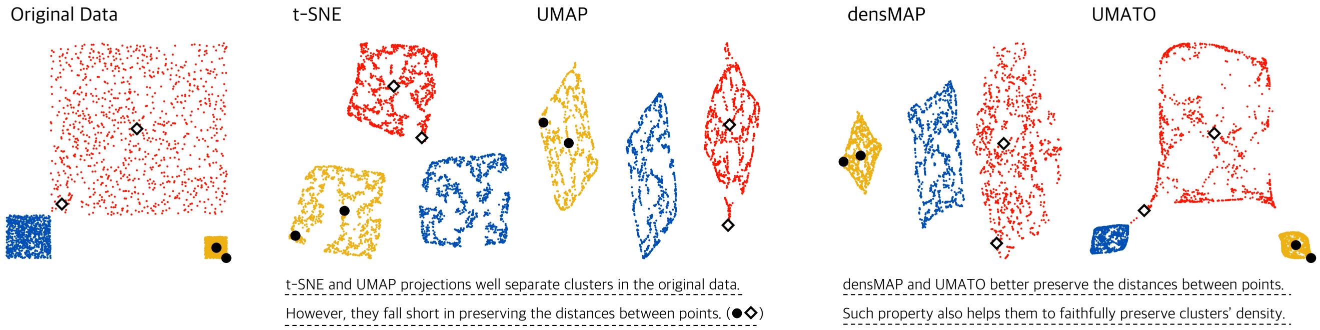

Figure 1: *Comparison of t-SNE, UMAP, densMAP [54], and UMATO [31] projections of a 2D dataset.* Although *t*-SNE and UMAP do not faithfully represent cluster density or distances between data points, they are often misused to analyze such structures. Our research investigates why this misuse happens and explores strategies to address it. [This figure is interactive in Adobe Acrobat reader, where the underlined texts can be clicked]

2.1 How *t*-SNE and UMAP Impacts Visual Analytics

We first explain DR, *t*-SNE, and UMAP. Then, we explain how the two techniques influence visual analytics.

Dimensionality reduction. DR, e.g., *t*-SNE, UMAP, PCA [57], is an essential tool for visually analyzing high-dimensional data [18, 23, 34, 42, 56]. These techniques receive a high-dimensional dataset as an input and produce a 2D representation that preserves important characteristics (e.g., local neighborhood structure or distances between clusters) of the original data. Using DR, any high-dimensional data can be visualized using a single scatterplot. Such effectiveness makes DR widely used for visual analytics across diverse domains [8], including bioinformatics [14, 70], machine learning [27, 38], and finance [12].

***t*-SNE.** Since its first introduction in 2008 [66] as a variant of SNE [26], *t*-SNE has become one of the most widely used DR techniques. It projects a given high-dimensional data into a low-dimensional space by minimizing the divergence between two distributions: one representing pairwise similarities of the points in the original space and the other in the low-dimensional space. By leveraging Student's t-distribution to model similarities in the low-dimensional space, *t*-SNE significantly improves SNE in accurately capturing the local neighborhood structure of the input data [45].

UMAP. This technique is introduced in 2018 [48] and has quickly gained popularity in diverse fields, including visual analytics. UMAP captures the local structure of high-dimensional data by constructing a *k*-nearest neighbors (*k*NN) graph of the data. It then optimizes a low-dimensional representation by minimizing the cross-entropy between the fuzzy topological representations of *k*NN graphs in the high and low-dimensional spaces.

Impact of *t*-SNE and UMAP in visual analytics. These two techniques significantly influence the visualization and visual analytics fields by motivating and grounding numerous follow-up studies. At first, these techniques are frequently utilized in visual analytics systems [11, 47, 75]. These systems typically use projections to provide an overview of the data, allowing users to select a subset of data through interactions such as brushing. The systems then enable detailed data exploration of the data subset using auxiliary visualizations [22, 43, 64]. The visualization community also attempts to accelerate these techniques using GPU [58, 65] or progressive algorithms [37, 40] to make them more responsive. Furthermore, the community improves the faithfulness of *t*-SNE and UMAP in representing original high-dimensional data, e.g., by proposing new variants [31, 50] or faithfulness metrics [30, 77].

Our contribution. Motivated by the widespread use of *t*-SNE and UMAP in visual analytics, we aim to investigate and improve how people use these techniques in practice. From our literature review and interview studies, we identify the (1) common misuse of these techniques within visual analytics, (2) the factors contributing to this misuse, and (3) why previous approaches fall short in mitigating the misuse. We propose future directions that promote the appropriate use of DR as a norm.

2.2 Previous Efforts to Address the Misuse

We identify literature from the visualization, machine learning, and bioinformatics domains that contributes to addressing the misuse of *t*-SNE and UMAP. These studies mostly conduct quantitative experiments that compare the utility of different DR techniques in supporting diverse analytic tasks. They also introduce a new DR technique that overcomes the limitations of *t*-SNE and UMAP. Another branch aims to inform the public about the weaknesses of *t*-SNE and UMAP, providing guidelines to avoid the inappropriate usage of these techniques.

Quantitative experiments. These works execute experiments to compare the suitability of the projections generated by diverse DR techniques on different analytic tasks. For example, Xia et al. [74] conduct a user study to test the effectiveness of DR techniques in supporting cluster analysis. They reveal that *t*-SNE and UMAP are most effective for cluster identification tasks but poorly support tasks like density comparison or distance comparison. Ventocilla and Reveiro [69] investigate the alignment between human task accuracy of cluster analysis and clustering metrics, leading to a similar conclusion to Xia et al. [74]. Jeon et al. [33] conduct a study using their newly proposed quality metrics for DR, showing that *t*-SNE and UMAP work poorly for investigating cluster density or separability. Aligned with these results, Lause et al. [44] also show the inappropriateness of *t*-SNE and UMAP in investigating global cluster arrangement, focusing on single-cell RNA data.

Advanced DR techniques. Diverse research proposes new DR techniques as alternatives to *t*-SNE and UMAP that mitigate their weaknesses in accurately representing the original high-dimensional data. Narayan et al. [54] verify that *t*-SNE and UMAP poorly represent cluster density and propose den-SNE and densMAP as alternatives. Trimap [1], PacMAP [71], and UMATO [31] improve UMAP in terms of capturing the global structure (e.g., pairwise distances between data points) of the original data. Global *t*-SNE [78] improves *t*-SNE in the same direction. Fig. 1 illustrates the superiority of alternative techniques against *t*-SNE and UMAP in preserving cluster density and distances between data points.

Guidelines for proper use of *t*-SNE and UMAP. These studies inform the limitations of *t*-SNE and UMAP to the public, guiding practitioners to use these techniques more appropriately. Wattenberg et al. [72] provide guidance on appropriately using *t*-SNE, and Coenen and Pearce [16] offer similar insights for UMAP. These works caution practitioners against relying on the global structure presented by *t*-SNE and UMAP and emphasize the substantial impact of hyperparameter selection on the faithfulness of resulting projections. Kobak et al. [41] show that initialization severely affects the faithfulness of the resulting projection and recommend using PCA for initializing *t*-SNE and UMAP. Chari and Pachter [9] demonstrate the case to which *t*-SNE and UMAP can lead to unreliable exploratory analysis in bioinformatics.

Our contribution. These studies offer actionable guidance to properly use *t*-SNE and UMAP along with the evidence that these techniques are often susceptible to misuse. However, we reveal that despite these efforts, both techniques are persistently misused. We identify that this is because practitioners lack motivation to engage with the literature, often due to the difficulty of reading papers. We therefore propose automating the process of finding optimal projections for analysis as a potential solution.

3 Research Objectives

We aim to contribute to addressing the misuse of *t*-SNE and UMAP. We set four research objectives to achieve this goal.

O1 **Verify the misuse of *t*-SNE and UMAP.** We want to find evidence that *t*-SNE and UMAP are widely used in visual analytics and are more frequently misused than alternative techniques. We thus want to provide a rationale for focusing on these two techniques.

O2 **Understand why practitioners misuse these techniques.** We aim to investigate the underlying cause of the misuse. This investigation grounds our suggestions for future directions to mitigate the misuse (O4).

O3 **Understand why previous efforts fall short in addressing the misuse.** We investigate why existing efforts have failed to prevent the persistent misuse of t-SNE and UMAP. As with O2, this investigation supports our suggestion of a new strategy to overcome the limitations of previous approaches (O4).

O4 **Introduce *future strategies* to promote the appropriate use of DR.** Finally, based on our findings, we want to introduce new directions and action items to mitigate the misuse of *t*-SNE and UMAP. We thereby contribute to enhancing the overall reliability of DR-based visual analytics.

The remaining parts of this paper are dedicated to achieving these objectives. First, we conduct a literature review (Sect. 4) on visual analytics papers using DR to investigate the extent to which *t*-SNE and UMAP are misused (O1) We deepen our investigation on misuse through an interview study with practitioners who regularly use DR for their research (Sect. 5), observing the underlying causes of the misuse (O2). We then interview expert researchers who study DR (Sect. 6) to obtain insights on why previous approaches are not effective to mitigate the misuse (O3). Based on the findings, we suggest future directions to remedy the misuse of *t*-SNE and UMAP in visual analytics (Sect. 7; O4).

4 Literature Review

We execute a literature review to confirm that *t*-SNE and UMAP are commonly misused in visual analytics (O1). We want to confirm that (**H1**) *t*-SNE and UMAP dominate the use of DR techniques in visual analytics, providing a rationale on why we focus on two techniques. We also aim to hypothesize the existence of two types of misuse: (**H2**) the use of *t*-SNE and UMAP for unsuitable tasks, and (**H3**) the lack of appropriate justifications. In the following sections, we first explain our protocol (Sect. 4.1). Then, in Sect. 4.2, we detail the results of the paper review, including common analytic tasks for DR and the reasoning behind the selection of DR techniques. In Sect. 4.3, we examine the suitability of widely used DR techniques for analytic tasks. Finally, in Sect. 4.4, we discuss our results that confirm the hypotheses. We discuss our takeaways in Sect. 4.5.

4.1 Protocol

The review consists of four steps: paper search, categorization, task suitability review, and quantitative analysis.

Paper Search. Our primary goal is to investigate the misuse of *t*-SNE and UMAP in visual analytics. Hence, we search for *papers that propose a visual analytics approach, framework, or systems* incorporating DR. We query papers that satisfy two conditions: (1) the term "`visual analytics`" or "`visual analysis`" appears in the title or abstract, and (2) one or more of the terms between "`dimensionality reduction`", "`dimension reduction`", "`multidimensional projection`", and "`multidimensional scaling`" are mentioned in the full text. We use IEEE Xplore and Wiley online library to search papers published in major data visualization journals and conferences (e.g., IEEE VIS, TVCG, CG&A, PacificVis, EuroVis, CGF, IVIS). We filter out papers published before 2008, the year *t*-SNE is announced. Then, we inspect each paper and remove the papers that do not stay within our search scope from our list. For example, papers that introduce novel DR techniques [43], interaction techniques [29], or visual analytics techniques that can be applied to any kind of scatterplot [36] are excluded. The list of identified papers can be found in Appendix A.

Categorization. We categorize the identified papers according to three criteria: (1) the DR techniques used, (2) the target analytic tasks supported by DR projections, and (3) the reasoning behind the selection of specific DR techniques. We select the first two criteria to investigate the extent to which DR techniques are misused for tasks that are not suitable for them (H2). We select reasonings as an additional criterion to check whether they justify the use of *t*-SNE, UMAP, and other DR techniques in an inappropriate way (H3). We follow the thematic coding process for the categorization. To begin with, two coders interdependently categorize the papers. Then, they merge and revise their categorizations through three iterative discussions (Sect. 4.2).

Task suitability review. We identify the suitability of DR techniques for analytic tasks by examining research that verifies the weaknesses and strengths of different DR techniques [31, 33, 54, 74] (Sect. 2.2). We depict the results in Sect. 4.3.

Quantitative Analysis. We quantitatively analyze the categorized papers to verify H1 and H2. Detailed statistical results are presented in Sect. 4.4.

4.2 Paper Search and Categorization Results

We retrieve 312 papers from online libraries in total. After screening and filtering, we retain 136 papers. The categories we identify from them are described below.

4.2.1 DR Techniques. We identify 18 DR techniques in total. Among them, we find four commonly-used techniques (*t*-SNE, UMAP, PCA, and MDS) where the usage of each technique is more than 20 times. The other 14 techniques are used fewer than five times each, and we categorize them all under "others."

4.2.2 Analytic Tasks. Our review reveals that analytic tasks using DR can be divided into seven categories (Fig. 2). Detailed descriptions of each task are as follows.

Neighborhood identification. This task aims to find data points similar to a target point based on the proximity within the projection. Since neighborhood identification supports many other analytic tasks, such as cluster identification, preserving local neighborhood structure is often considered the most important criterion when evaluating DR projections [31, 51, 56].

Outlier identification. This task is about identifying outliers within projections. Analysts often count the number of outliers in the data [20] or determine whether a point is a cluster member or outlier [74]. The task is typically leveraged for examining the quality of class labelings by identifying points that have high uncertainty about their class membership [28].

Cluster identification. This task involves identifying clusters within DR projections. Analysts typically count the number of clusters [20] or label clusters interactively using selection tools such as lasso or box-shaped brushes [15, 49, 74]. This task often includes investigating subclusters within existing clusters [20]. Visual analytics systems often provide auxiliary visualizations to show detailed information about the identified clusters [10, 47].

Point distance investigation. Similar to the cluster distance investigation task, this task investigates the distance between data points as a proxy for their high-dimensional dissimilarity. It can be interpreted as a "continuous" version of the neighborhood identification task.

Table 1: *The coverage of analytic tasks using DR (rows) by references (columns).* We represent the references using the last name of the first author. (Xia: Xia et al. [74], Ete.: Etemadpour et al. [20], Bre.: Brehmer et al. [7], Non.: Nonato and Aupetit [56], Sed.: Sedlmair et al. [61]), Cas. : Cashman et al. [8]

Task	Xia	Ete.	Bre.	Non.	Sed.	Cas.
Neighborhood Identification		•		•		•
Outlier Identification	•	•		•		•
Cluster Identification	•	•	•	•	•	•
Point Distance Investigation				•		•
Class Separability Investigation			•	•	•	•
Cluster Distance Investigation	•	•	•	•	•	•
Cluster Density Investigation	•	•		•		

Class separability investigation. This task investigates how distinctly different classes are separated in the projections, where the classes are color-coded. The task is commonly performed when DR techniques are employed to explain the behavior of a supervised machine learning model, particularly to illustrate how the model distinguishes between different classes [59, 62].

Cluster distance investigation. This task uses the distance between well-separated clusters as a proxy for their similarity in the original high-dimensional space. The clusters can be explicitly labeled (i.e., color-coded) or implicitly represented by data distribution [61].

Cluster density investigation. This task identifies and compares the density of clusters using cluster density as a proxy for the variability of data points within each cluster [54].

Task coverage validation. To validate the comprehensiveness of our categorization, we examine whether the analytic tasks in our list are included in prior task taxonomies within the visualization field. We review Etemadpour et al. [20], Xia et al. [74], Brehmer et al. [7], Nonato and Aupetit [56], Sedlmair et al. [61], and Cashman et al. [8]. We find that all tasks are covered by at least two previous studies (Table 1), confirming that our categorization covers all important tasks in DR.

4.2.3 Reasonings. We identify seven large categories of reasoning used to justify the selection of DR techniques. We define each reasoning in Table 2. It is worth noting that a substantial amount of papers (44%; Fig. 6) do not mention specific reasoning. We additionally categorize these papers as "No reason".

Faithfulness. Researchers justify the usage of DR techniques based on their faithfulness, or the capability to accurately represent the original structure of the high-dimensional data without distortions. This reasoning mostly relies on references to benchmark studies that examine the faithfulness of DR techniques [19, 74]. One notable finding is that researchers often cite the original papers introducing DR techniques to support claims about their capability to preserve global structure, e.g., the original UMAP paper [48], which is not

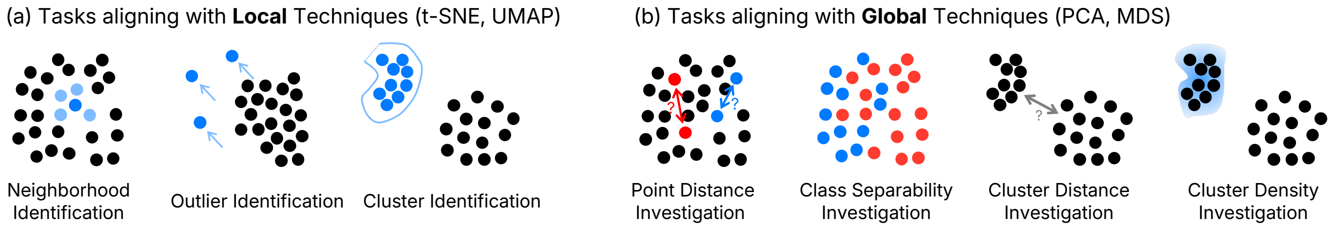

Figure 2: *Illustrations of the analytic tasks using DR and their alignment to local and global DR techniques.* Our literature review identifies seven types of analytic tasks using DR. *t*-SNE and UMAP are suitable for neighborhood identification, outlier identification, and cluster identification tasks but inappropriate for other tasks.

Table 2: *The definition of the reasonings that we identify from the literature review (Sect. 4.2.3)* Except for extensibility and simplicity, all reasonings are leveraged to justify the use of *t*-SNE and UMAP.

Reasoning	Definition
Faithfulness	The degree to which DR techniques accurately represent the original structure of the high-dimensional data without distortions.
Popularity	The degree to which DR techniques are widely known and used by practitioners in the visual analytics field.
Scalability	Computational efficiency in executing DR techniques.
Interpretability	The degree to which DR techniques yield visually distinct, analyzable clusters, enabling clear explanation of the data.
Stability	The degree to which DR techniques produce projections that are stable against hyperparameter change or the stochastic nature of DR.
Extensibility	The degree to which DR techniques can be adapted or expanded to accommodate diverse data conditions or input formats.
Simplicity	The degree to which practitioners can readily understand and apply DR techniques.

always correct [16, 31, 36, 54, 74]. We also find that several papers claim the faithfulness of UMAP without references.

Popularity. Researchers justify the use of DR techniques based on their popularity, which indicates the degree to which the techniques are widely acquainted and used by practitioners in the visual analytics field. For example, papers mention employing *t*-SNE because it is a "default option" in visualizing high-dimensional data or is "widely recognized" by the research community. These papers also highlight specific research communities, such as biology, computer vision, and document clustering, where these techniques are commonly used [8].

Scalability. The use of DR techniques is also justified by their computational efficiency, which enhances the responsiveness of visual analytics systems. For example, papers state that efficient GPU implementations [55, 58] facilitate the effectiveness of DR techniques.

Interpretability. Researchers use DR techniques because they enable a clear explanation of the data with projections that contain visually distinct and, thus, analyzable clusters. This finding aligns with the work by Morariu et al. [52] and Doh et al. [17], in that they also identify clear cluster separation as a key factor influencing the preference of DR projections.

Stability. Researchers justify the selection of DR techniques by highlighting their stability (i.e., the degree to which DR techniques produce projections that are stable against hyperparameter change or the stochastic nature of DR) as a means to improve the reproducibility of their research. For example, one paper argues that *t*-SNE is stable due to its non-convex optimization [2].

Extensibility. We find that researchers justify the use of DR techniques based on their extensibility or their ability to adapt or expand to accommodate diverse data conditions or input formats. For example, some papers use DR techniques because they are parametric, i.e., support new data points to be dynamically projected after initial projection [60], particularly for analyzing streaming, online data.

Simplicity. A few papers mention that they select DR techniques that are simple and thus can be easily understood and applied by practitioners. Among the four major techniques, only PCA has been explicitly justified by this reasoning.

4.3 Suitability of DR Techniques to Analytic Tasks

We assess the suitability of four major DR techniques (*t*-SNE, UMAP, PCA, and MDS) to the analytic tasks identified in Sect. 4.2.2. This is done by revisiting previous studies that evaluate DR techniques and analyze the alignment between the DR techniques and analytic tasks (Sect. 2.2). This analysis helps examine whether researchers are applying DR to tasks that are suitable (H1). Here, we define that a DR technique is suitable for an analytic task if it preserves the structural characteristics corresponding to that task. This means that analytic tasks can be reliably conducted when suitable DR techniques are used.

4.3.1 Tasks Suitable for t-SNE and UMAP. *t*-SNE and UMAP focus on preserving local neighborhoods by positioning neighboring points close together in the projection while separating non-neighboring points. They are thus commonly referred to as local techniques. Several studies demonstrate that they show state-of-the-art performance in preserving local structures, both empirically

[19, 31, 51] and theoretically [45]. We categorize the following tasks that are suitable for *t*-SNE and UMAP.

Neighborhood identification task is more suitable for *t*-SNE and UMAP. As aforementioned, *t*-SNE and UMAP directly aim to preserve local neighborhood structure. This makes them better suited for neighborhood identification tasks than alternative techniques by design [1, 33, 68, 78], which also have been empirically validated [31, 51].

Outlier identification task is more suitable for *t*-SNE and UMAP. Since projections generated by local techniques clearly distinguish neighboring and non-neighboring points, they can effectively separate outliers from clusters. Xia et al. [74] empirically validate that *t*-SNE and UMAP are the most effective DR techniques in supporting the outlier identification task, outperforming alternative techniques like PCA.

Cluster identification task is more suitable for *t*-SNE and UMAP. As *t*-SNE and UMAP locate neighboring points close and non-neighboring points far away [31, 51, 68], they clearly represent individual high-dimensional clusters as 2D clusters, thus suitable for the cluster identification task. The recent user study by Xia et al. [74] shows that the participants perform best when identifying clusters with *t*-SNE and UMAP projections.

4.3.2 Tasks Suitable for PCA and MDS. PCA and MDS are DR techniques that are more effective in preserving global pairwise distances between data points compared to local techniques like *t*-SNE and UMAP [31, 56, 63, 67, 74]. They are usually referred to as global techniques. The following tasks are suitable for these DR techniques.

Point distance investigation task is more suitable for PCA and MDS. These techniques are designed to directly preserve the pairwise distance structures more effectively compared to local techniques. They are thus more suitable than *t*-SNE and UMAP in investigating distances between data points (Fig. 1 dots and diamonds). Several studies [1, 31, 51] propose techniques that improve UMAP in preserving global pairwise distances between points, such as UMATO [31] and TriMap [1].

Class separability investigation task is more suitable for PCA and MDS. The superiority of PCA and MDS in preserving distances between data points makes them more precisely exhibit the separability between class labels [33, 72, 73]. For example, *t*-SNE and UMAP are widely reported to exaggerate class separation compared to other techniques [3, 5, 6, 33]. Wattenberg et al. [72] show that hyperparameter choices can significantly distort the global relationship between classes in *t*-SNE, including their separability.

Cluster distance investigation task is more suitable for PCA and MDS. PCA and MDS better preserve pairwise distances between points within each cluster compared to alternatives, making the inter-cluster distances in the resulting projections meaningful. These techniques are thus suitable for tasks involving cluster distance analysis. Many computational benchmarks have validated the superiority of PCA and MDS in supporting the cluster distance investigation task [31, 33, 51, 69, 71]. This implies the appropriateness of these techniques for supporting cluster distance investigation. Xia et al. [74] empirically show that global techniques like PCA enable users to perform this task more accurately than local techniques. In contrast, Wattenberg et al. [72] and Coenen et al. [16] also inform that the distance between clusters lacks meaning in *t*-SNE and UMAP projections, respectively.

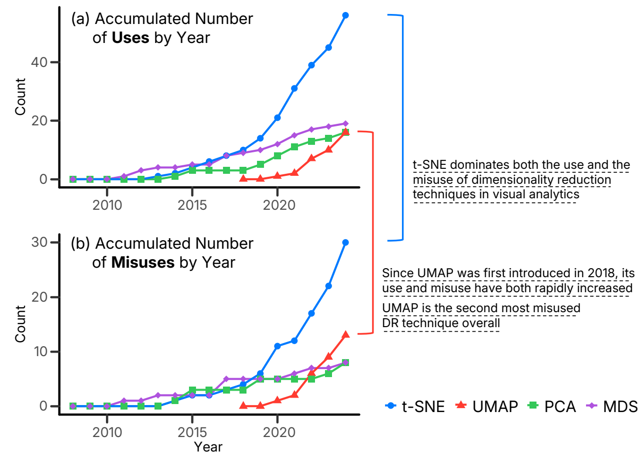

Figure 3: *The trend of accumulated number of papers that use (a) or misuse (b) of four major DR techniques.* We collect papers published from 2008, the year *t*-SNE is introduced. Note that UMAP's data also starts from the year it is released (2018). [This figure is interactive in Adobe Acrobat reader, where the underlined texts can be clicked]

Cluster density investigation task is more suitable for PCA and MDS. These techniques depict the similarity between data points as low-dimensional proximity and thus can more sensitively depict the differences in cluster densities. In contrast, local techniques like *t*-SNE and UMAP poorly reflect their true similarity in high-dimensional space [1, 31, 54] as they only focus on neighboring points. As a result, *t*-SNE and UMAP projections poorly represent cluster density (Fig. 1), which motivates the development of improved techniques such as den-SNE and densMAP [54]. The superiority of global techniques in supporting the density investigation task is also validated by Xia et al. [74] through user studies and Jeon et al. [33] via case studies.

4.3.3 Validity of the Suitability Analysis. One notable finding in our suitability analysis is that *t*-SNE and UMAP perform better on all "identification" tasks while PCA and MDS excel at "investigation" tasks. This result aligns with the fundamental differences in how these methods interpret distances. *t*-SNE and UMAP prioritize preserving local neighborhood structures by effectively treating similarity as a binary function (neighbors or non-neighbors), making them well-suited for tasks that require identifying distinct clusters or groups. In contrast, PCA and MDS interpret distances as continuous values, enabling more accurate interpretation of relative distances between points. This finding supports the validity of our task categorization in capturing the alignment between DR techniques and the tasks that are suitable to.

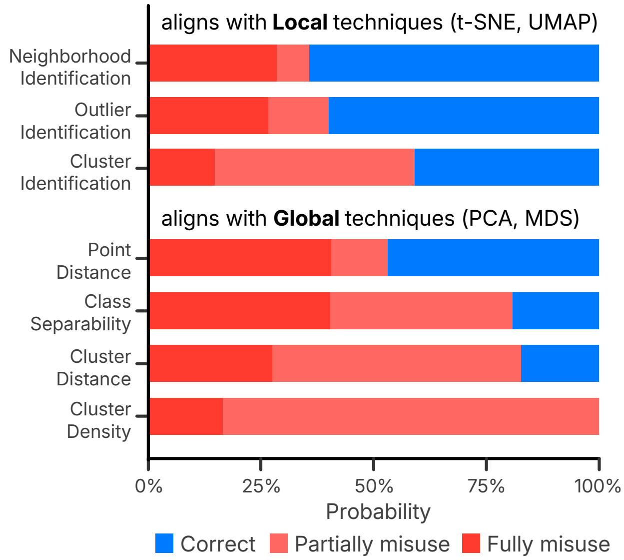

Figure 4: *The ratio of appropriate use and misuse of DR techniques by analytic tasks.* DR is appropriately used for tasks that align with local techniques (top 3) but not for those that align with global techniques (bottom 4). This result indicates that local techniques (e.g., *t*-SNE, UMAP) are overtrusted even for tasks that are not suitable. Papers are marked as "fully misused" if all tasks targeted by the paper are not supported by the employed DR technique. Papers with partial support are marked as "partially misused."

4.4 Findings

We present the findings from the quantitative analysis of the identified papers (Sect. 4.2). We reveal that *t*-SNE and UMAP are the most commonly adopted DR techniques (H1). Yet, researchers often utilize them across any tasks, making them simultaneously the most commonly misused DR techniques (H2). We also observe that many papers leverage *t*-SNE and UMAP without justifications or with improper reasonings (H3).

(Finding 1) *t*-SNE and UMAP dominate the use of DR. While the number of visual analytics papers using DR has increased over the years, this growth is largely driven by *t*-SNE and UMAP (Fig. 3; H1). *t*-SNE, for example, appears in more than half of the identified papers (75 out of 136), more than twice as often as the runner-up. UMAP is used in 31 papers. However, UMAP's adoption is increasing at a much steeper rate compared to PCA and MDS, enabling it to achieve parity with these established techniques within only six years. These findings highlight that misusing *t*-SNE and UMAP can have a more harmful impact than misusing other techniques.

(Finding 2) *t*-SNE and UMAP are used for *any* tasks. We identify that t-SNE and UMAP are commonly used for both suitable and unsuitable tasks, confirming H2. For each analytic task, we compute the proportion of papers that employ suitable DR techniques relative to all papers addressing the task. As a result (Fig. 4), we find that tasks supported by local techniques have a high proportion

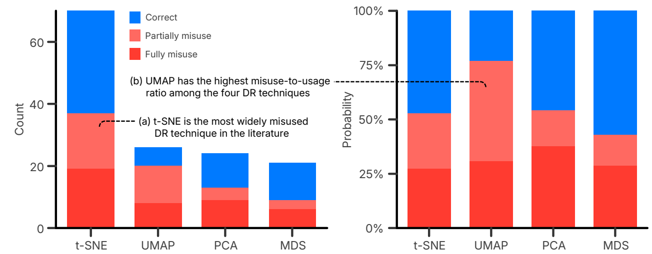

Figure 5: *The number of appropriate use and misuse of DR by techniques (left) and their ratio (right).* As with Fig. 4, papers are marked as "fully misused" and "partially misused" if all tasks targeted by the paper are entirely or partially not supported by the used DR techniques. The analysis reveals that *t*-SNE and UMAP dominate the misuse of DR.

of proper usage. However, tasks requiring global techniques have a substantially lower rate of proper usage. This indicates that researchers appropriately employ local techniques (*t*-SNE and UMAP) when required, while also correctly avoiding global techniques for these tasks. However, tasks requiring global techniques have a substantially lower rate of proper usage, indicating that *t*-SNE and UMAP are misused even for unsuitable tasks.

This tendency makes *t*-SNE and UMAP to dominate the misuse of DR techniques in practice. We find that *t*-SNE is the most widely misused DR technique in our list of papers (Fig. 5a). Regarding UMAP, we observe that its misuse has recently increased rapidly, making it the runner-up (Fig. 3b). We also find that UMAP has the highest misuse-to-usage ratio among the four major DR techniques (Fig. 5b). In summary, *t*-SNE and UMAP are misapplied more frequently than other techniques, underscoring the need for increased efforts to address these misuses.

In addition, we observe that a similar pattern persists when we exclude papers published before 2016 and 2019—the years when two influential papers guiding researchers on the use of *t*-SNE and UMAP are released [16, 72] (Appendix C). This indicates that the misuse of *t*-SNE and UMAP persists despite existing efforts to inform practitioners how to use these techniques properly.

(Finding 3) *t*-SNE and UMAP are used without reasonings or with improper reasonings. We find that more than 40% of papers do not explicitly justify their choice of DR techniques (Fig. 6; H3), and this trend persists for *t*-SNE and UMAP. These papers often discuss the general need for DR or describe the techniques' characteristics rather than explaining why specific techniques are chosen. This result implies that practitioners may perceive DR technique selection—including the case of *t*-SNE and UMAP—as requiring less critical evaluation, suggesting a lack of clear understanding of the appropriate way of using DR. Our subsequent interviews (Sect. 5) further reinforce this observation in the context of *t*-SNE and UMAP usage.

We also identify that faithfulness is more widely used to justify the use of *t*-SNE and UMAP compared to PCA and MDS, but is referred to even when these techniques are misused (Fig. 6). This result thus supports our implications that researchers often lack proper understanding of DR; they may not know that *t*-SNE and

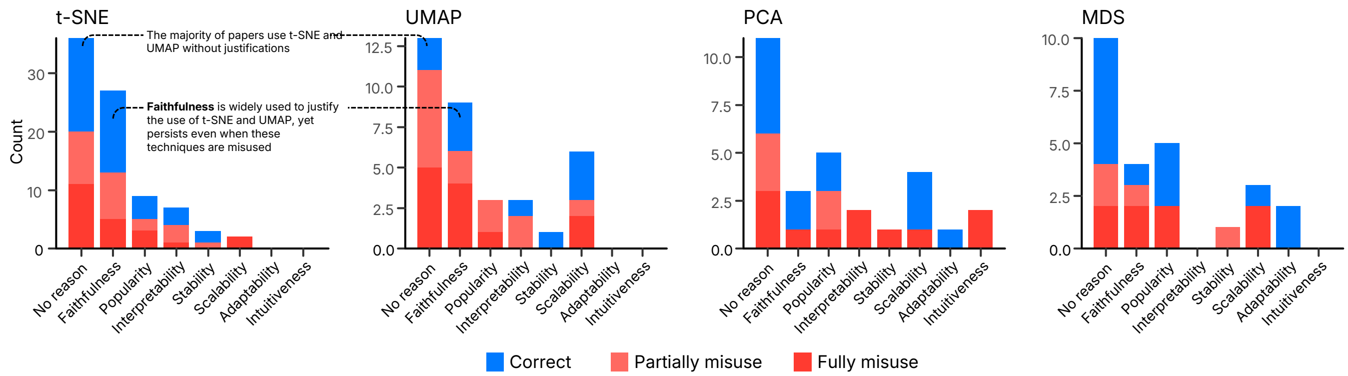

Figure 6: *The number of appropriate use and misuse of DR techniques by reasonings.* The reasonings (x axis) are sorted in descending order based on the number they are referenced to justify the use of *t*-SNE (leftmost chart).

UMAP are faithful in preserving local structure but not in global structure (Sect. 4.3).

4.5 Takeaways

The following are key takeaways from our literature review.

Misuse of *t*-SNE and UMAP exists. We quantitatively verify that *t*-SNE and UMAP are commonly misused in practice (Findings 2 and 3).

Researchers may have a limited understanding on how to use DR properly. We find that researchers often use *t*-SNE and UMAP without proper justification or regard them as faithful for unsuitable tasks (Finding 3), implying limited awareness of various DR techniques and their suitability to analytic tasks. Indeed, the misuse of *t*-SNE and UMAP itself suggests that practitioners do not know how to properly use DR. Our interview study (Sect. 5) reaffirms this observation.

Researchers may lack the motivation to use *t*-SNE and UMAP properly. Our review indicates that papers misusing *t*-SNE and UMAP have passed peer review and been published in major visualization conferences and journals. This finding suggests that reviewers frequently overlook the importance of using DR appropriately. Furthermore, the frequent absence of clear justifications for selecting DR techniques (Finding 3) implies that researchers often do not recognize the necessity of using DR properly.

5 Interview Study with Practitioners

We want to investigate why practitioners misuse *t*-SNE and UMAP in practice (O2). To this end, we conduct interviews with practitioners who regularly use DR techniques for their research or visual analytics work. In the following sections, we first explain our study design (Sect. 5.1). Then, in Sect. 5.2, we detail our findings. Finally, we discuss the takeaways in Sect. 5.3.

5.1 Study Design

Participants and recruitment. We want to diversify our participants in terms of their experience on DR. We first aim to achieve diversity in the domains in which participants work. To do so, we recruit both visual analytics researchers and domain researchers who have experience in visually analyzing and presenting their data using *t*-SNE and UMAP. For visual analytics researchers, we randomly select papers from our literature review (Sect. 4), prioritizing diversity in target data and problem domains. We then contact either the first or the corresponding author via email to increase diversity in participants' expertise and visualization literacy. For domain researchers, we ensure that they are from distinct disciplines without overlap. We recruit participants from a local university through an internal web community. We also employ snowball sampling [24] to expand our participant pool. In total, we interview 12 participants (six visual analytics researchers and six domain researchers) with diverse occupations and research experience (Table 3).

Interview protocol. We interview participants in a semi-structured manner. We first ask the participants to give consent for their participation. While doing so, we clarify that the paper identifies potential risks in the participants' reasoning and emphasize that our research is not meant to blame the participants or other researchers. We then ask participants a series of questions. The questions mainly ask the participants (1) their expertise in DR, *t*-SNE, and UMAP, (2) their experience and justifications in using *t*-SNE and UMAP, and (3) the difficulties that occur while using these techniques (our questionnaire is in Appendix B). The interviews are conducted via a recorded Zoom call, where we transcribe the interview using a commercial speech-to-text service. We compensate participants with the equivalent of 15 USD. All interviews are finished within 40 minutes.

5.2 Findings

(Finding 1) Practitioners have limited literacy on DR. We find that the participants have difficulties in properly understanding not only *t*-SNE and UMAP but also other DR techniques, aligned with our takeaways from the literature review (Sect. 4.5). For instance, five participants report difficulties in choosing the final projection for deployment, as the outcomes vary significantly depending on hyperparameter configurations. Three of these participants admit that they are unsure of the proper way to set these hyperparameters.

Table 3: *The demographics of the participants in our interview study with practitioners.* Our aim in recruiting participants is to maximize diversity in research fields and experience (exp.). VA stands for visual analytics. We chronologically order the participants based on the date of the interviews.

	Occupation	Age	Gender	Type	exp. in VA	exp. in DR	Domain
P1	Professor	35	Male	Visual Analytics Researcher	10 years	7 years	·
P2	Undergraduate Student	22	Female	Visual Analytics Researcher	2 years	1 year	·
P3	Research Scientist	30	Male	Domain Researcher	·	6 years	Computer Vision
P4	Research Scientist	30	Male	Domain Researcher	·	4 years	Signal processing
P5	Ph.D. Student	28	Female	Visual Analytics Researcher	6 years	5 years	·
P6	Ph.D. Student	28	Male	Domain Researcher	·	4 years	HCI
P7	Ph.D. Student	29	Male	Domain Researcher	·	4 years	Chemistry
P8	Ph.D. Student	24	Female	Visual Analytics Researcher	3 years	2 years	·
P9	Research Scientist	30	Male	Domain Researcher	·	7 years	NLP
P10	MS Student	22	Male	Visual Analytics Researcher	2 years	1 year	·
P11	Staff Engineer	34	Male	Visual Analytics Researcher	2 years	2 years	·
P12	Postdoctoral Researcher	34	Male	Domain Researcher	·	4 years	Bioinformatics

Second, participants often do not have sufficient knowledge on alternative DR techniques. For example, four participants say they are unaware of methods other than *t*-SNE, UMAP, and PCA, and two of these participants mention that they can hardly tell the differences among these three techniques.

We also noticed that practitioners' limited literacy make them "immune" to using *t*-SNE and UMAP for their research. Participants mention that the popularity of *t*-SNE and UMAP makes them less likely to invite criticism of their analytic results or systems. For example, three participants mention using UMAP because they feel that using alternative techniques may subject their paper to criticism from reviewers. One participant notes that UMAP is a *safe* technique because it is the state-of-the-art technique. This finding aligns with our observation that many papers we investigate do not provide appropriate reasoning for using *t*-SNE and UMAP (Sect. 4.4).

(Finding 2) Practitioners receive misleading suggestions. We find that the misuse also occurs because practitioners regularly receive potentially misleading suggestions to use *t*-SNE and UMAP. We identify three primary sources:

Fellow researchers. Five participants indicate that their fellow researchers recommend using *t*-SNE and UMAP. Two participants, in particular, mention that they unreservedly follow recommendations from their advisors or seniors. For example, a participant comments: *"My advisor recommended UMAP, and I used it without verification"*.

Research papers. Two domain researchers state that they used *t*-SNE and UMAP after they frequently encountered them in research papers in their domains (bioinformatics and chemistry). One participant, for instance, mentions that he regularly uses UMAP because it is frequently cited in recent publications in his domain (bioinformatics) for the same purpose he intends to use it for. This finding resonates with the observation of Cashman et al. [8] that these domains predominantly adopt *t*-SNE and UMAP over alternative DR techniques.

Language models. Two participants mention that they ask large language models (LLMs) to recommend a DR technique to use, where the models suggest *t*-SNE, UMAP, and PCA. Both participants report using ChatGPT for this purpose.

These suggestions can be misleading as they typically lack grounded evidence. For example, research papers often misuse *t*-SNE and UMAP (Sect. 4), so relying on these papers can lead to erroneous applications of these techniques. Moreover, as LLMs are trained on massive text corpora, their frequent recommendations of *t*-SNE and UMAP can be interpreted as reinforcing practitioners' bias toward these techniques based on their popularity. Despite the existence of materials informing the proper way of using DR (Sect. 2.2), practitioners' reliance on such misleading suggestions indicates their insufficient motivation to engage with these resources.

(Finding 3) Practitioners often cherry-pick hyperparameters. We observe one more misuse pattern: the cherry-picking of hyperparameters. Eight out of 12 participants report that they have experience in manually tuning the hyperparameters of *t*-SNE and UMAP. They report that they want to achieve either an interpretable or aesthetically pleasing projection with well-separated clusters. Four participants explicitly mention that they tune hyperparameters without understanding their effect on projection results.

5.3 Takeaways

We identify three key takeaways from the interview study with practitioners.

Practitioners lack understanding of how to use DR properly. The interview study clearly confirms that practitioners have a limited understanding of DR. Many participants (1) hold erroneous beliefs about the faithfulness of *t*-SNE and UMAP (Finding 1), (2) do not know how to select appropriate techniques (Finding 1), and (3) do not know how to properly set hyperparameters (Finding 3).

Practitioners lack motivation to use DR properly. The interview study suggests that practitioners are unaware of how to use DR effectively and lack motivation to do so, aligning with our third takeaway from the literature review (Sect. 4.5). For example, the

Table 4: *The demographics of the participants in our interview study with DR experts.* Our aim in recruiting participants is to maximize diversity in terms of research domain and seniority while maintaining their expertise level. Exp. and Pub. denote the years of research experience and the number of publications related to DR, respectively.

	Occupation	Age	Gender	Exp.	Pub.	Subdiscipline	Venue of DR-related papers
P1	Associate Professor	34	Male	12	>10	DR Faithfulness, Scalability	TVCG, VIS, EuroVis
P2	Associate Professor	48	Male	20	>10	DR Algorithm, Evaluation	TVCG, VIS, CGF, EuroVis, C&G
P3	Assistant Professor	39	Male	13	>10	DR Interpretation	TVCG, VIS, CHI, CGF
P4	Associate Professor	40	Male	20	>10	DR Interpretability, Evaluation	TVCG, VIS, CGF, C&G
P5	Assistant Professor	36	Male	11	4	Practical use of DR	TVCG, VIS
P6	Ph.D. Student	26	Male	4	3	DR Stability, Faithfulness	TVCG, VIS
P7	Ph.D. Student	26	Female	3	3	DR Faithfulness	EuroVis, CGF
P8	Ph.D. Student	28	Male	6	8	Visual analytics for DR	TVCG, VIS, CHI

perception of *t*-SNE and UMAP as "immune to criticism" arises from reviewers' insufficient interest in the proper use of DR. The limited understanding of DR (Findings 1 and 3) also suggests that practitioners are often unconcerned with this issue.

6 Interview Study with DR Experts

We want to understand why previous efforts are not effective in addressing the misuse (O3), grounding our new suggestions to remedy the problem (Sect. 7). To this end, we interview visualization researchers who study DR as their main research topic. We first discuss our study design in Sect. 6.1, then detail our findings and takeaways in Sect. 6.2 and Sect. 6.3, respectively.

6.1 Study Design

Participants and Recruitment. We establish two objectives in recruiting the expert researchers. First, we want our experts to have a sufficient **expertise** on (1) the underlying mechanism of DR and (2) how it is used in practice for visual analytics. Experts without the former may struggle to understand essential concepts to understand our problem, including the rationale behind different DR techniques and their alignment to analytical tasks. Conversely, experts without the latter may provide limited insights into addressing real-world misuse scenarios. Second, we want our pool of experts to be sufficiently **diverse** in terms of their experience on DR.

To achieve these goals, we recruit experts by randomly sending emails to the authors of the papers that contribute to addressing the misuse of DR, which we list in our related work section (Sect. 2.2. Note that we contact one of the corresponding authors and the first author to further diversify expertise and research experience. When recruitment is declined, we sample a new author from the same cluster and repeat the process. Table 4 depicts the demographics of our experts.

Interview protocol. We conduct an interview individually with each expert. One experimenter, the first author, manages all experiments. After the experts sign the consent form, the experimenter verbally describes the purpose of the experiment and how it will proceed. Then, we share three misuse patterns we find from the literature review (Sect. 4) and the interview with practitioners (Sect. 5): inappropriate selection of DR techniques, lack of proper justifications, and cherry-picking of hyperparameters. Subsequently, we ask interviewees to complete the following questionnaires, asking why the previous efforts in the literature to address the misuse have hardly achieved the goal.

- *"Why do you think these misuses persist despite the large body of literature that informs practitioners not to do?"*
- *"Why do you think practitioners are not motivated to read existing materials?"*

The expert answers each question verbally. We wrap up the interview by requesting the experts to share their thoughts on how the misuses can be addressed. We do not limit the interview duration to fully elicit the experts' insights, but all interviews end within 40 minutes. We compensate experts with the equivalent of 20 USD for their participation.

6.2 Findings

All experts agree that misuse persists despite existing efforts and share the following insights on underlying causes.

(Finding 1) Cultivating DR literacy is not easy. Experts note that cultivating DR literacy is not easy, even for trained researchers. Two experts especially mention that although survey papers exist to comprehensively inform the way of using DR properly, even these resources are difficult for novices to understand. One expert note that understanding these survey papers ultimately requires reading individual papers, which is a demanding task. This finding suggests that existing efforts to instruct practitioners—organizing and presenting information scattered in diverse papers [34] (Sect. 2.2)—hardly reduce the inherent difficulty of learning DR.

(Finding 2) Libraries are promoting the misuse of *t*-SNE and UMAP. Five experts note that highly polished and well-maintained libraries that serve *t*-SNE[1] and UMAP[2] intensify the misuse. They say that although practitioners want to test other DR techniques, they can hardly find and execute implementations, coming back to *t*-SNE and UMAP. Of these, three experts note the need for new libraries that serve diverse DR techniques. They mention that such libraries will help people be aware of diverse DR techniques and properly use them. P8 states: *"Good libraries, like scikit-learn, helped*

[1] https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html
[2] https://umap-learn.readthedocs.io/en/latest/

popularize t-SNE and UMAP. Similarly, if other techniques have good libraries, they will likely be used more often." Two experts also mention that not only DR techniques but also other infrastructures for utilizing DR, such as preprocessing or evaluation metrics, should be provided as libraries.

(Finding 3) Intrinsic bias of human perception is promoting the misuse of *t*-SNE and UMAP. Two experts causally mention the possibility that previous attempts fail because there exists an intrinsic bias towards well-separated clusters. They note that in such a case, it is natural for *t*-SNE and UMAP to be widely used because these techniques exaggerate cluster structure by design [32, 45]. This finding aligns with the recent finding of Doh et al. [17] that there exists a perceptual factor that makes practitioners prefer DR projections with clearly separated classes or clusters.

(Finding 4) Mitigating bias is an urgent problem. Three experts note that we should urgently address these misuses because they may compromise the reliability of scientific discoveries based on visual analytics. P5 states: *"False positives (due to the misuse of DR) are undoubtedly occurring somewhere at this very moment."* Two experts also mention that the urgency intensifies because misuses can propagate through citation networks; they emphasize that intervention is needed before these misuses become de facto standards. Both experts note that *t*-SNE and UMAP appear to have already approached this status, warning that mitigation will become increasingly difficult over time.

6.3 Takeaways

We should move beyond papers. Through the interviews, we identify the limitations of conventional efforts to mitigate bias, which are mostly based on academic papers (Finding 1). Moreover, experts emphasize the need for technical solutions to mitigate this bias (Finding 2). This suggests that we need to invest efforts that extend beyond traditional academic approaches to make proper use of DR as the norm.

We should act immediately. The interview suggests that immediate action is necessary to address the misuse of *t*-SNE and UMAP. Experts warn that the misuse may already have passed the point of no return (Finding 4).

7 Recommendation: Delegation to Machines

Based on our study findings, we cautiously but strongly suggest **delegating the configuration of DR for visual analytics to machines**. Imagine *VoyagerDR*, a hypothetical programming library with comprehensive knowledge of DR. This library is equipped with a function that, when given a high-dimensional dataset and a specified analytic task, automatically predicts and recommends ideal DR techniques and hyperparameters that maximize task performance. In theory, VoyagerDR would always guarantee the proper use of DR and enhance the reliability of visual analytics.

However, while appealing in concept, adopting VoyagerDR may be controversial because it may harm user agency. With VoyagerDR, users may heavily rely on this library and their understanding of DR will remain limited. This is problematic because finding an optimal DR projection is not solely about selecting the DR technique that best aligns with the target task; it also depends on factors such as available computational resources, time constraints, and the *practitioners' priorities*—such as whether faithfulness, stability, or efficiency takes precedence. For example, in time-sensitive scenarios, it might be more effective to bypass hyperparameter optimization in favor of faster system responsiveness. In this sense, even with such an idealized DR oracle, we should not abandon efforts to enhance practitioners' literacy as a path toward more reliable visual analytics.

We therefore call for investigating approaches that enable practitioners to benefit from Voyager while simultaneously cultivating their DR literacy. One way to achieve this is to make VoyagerDR more explainable. For example, enabling a `verbose` option by default to reveal how the library operates would be beneficial. Furthermore, creating an *Explainer* [39, 46], an interactive visualization that explains both VoyagerDR's operational processes and proper DR usage, will also be a plausible direction. Reflecting on the takeaways from the expert interviews, we call upon the community to pursue this direction—without delay (Sect. 6.3).

8 Discussions

Our discussion addresses three key questions regarding the reliable use of DR and other machine learning techniques.

8.1 Is VoyagerDR Feasible?

While VoyagerDR (Sect. 7), a DR oracle that automatically predicts optimal DR projections for a given dataset and task, may appear challenging to realize, we believe the necessary knowledge and technology already exist for its implementation. First, the visualization research community has accumulated substantial understanding about the mapping between techniques and suitable tasks through numerous benchmark studies [4, 20, 74] and guidelines for proper DR technique usage [16, 72] (Sect. 2.2). The primary challenge lies in formalizing these guidelines into a format that VoyagerDR can interpret—a task for which existing approaches like Draco [53] provide inspiration. Furthermore, selecting appropriate models or hyperparameters based on dataset characteristics is a widely studied problem in machine learning under the umbrella of AutoML [21]. Indeed, recent work by Jeon et al. [35] successfully predicts the most accurate technique by measuring dataset complexity. We argue that these approaches can be extended to encompass a broader range of techniques and, ultimately, the full hyperparameter space.

Given this feasibility, we envision the realization of VoyagerDR in the near future.

8.2 Is DR Misused in Other Fields?

This research reviews relevant literature in the field of visual analytics. By doing so, we reveal that visual analytics researchers often use *t*-SNE and UMAP inappropriately. However, while the same issue is present in other research fields (e.g., bioinformatics [9, 44, 54]), our findings may not effectively reach researchers in these areas due to our emphasis on visual analytics. Our interviews with domain researchers help address this gap (Sect. 5), but they cover only four areas: machine learning, HCI, chemistry, and bioinformatics. Examining the use of DR in various domains beyond visualization could help us convey our message to a wider audience and uncover more broadly applicable solutions. Cashman et al.

[8] recently present a seminal step in this direction, focusing on a limited set of domains (Physics, Chemistry, Biology, and Business). We encourage continued efforts to broaden the impact of research artifacts from the visualization community.

8.3 Are We Properly Using Machine Learning for Visual Analytics?

In this research, we find that practitioners frequently misuse *t*-SNE and UMAP, often considering these techniques as one-size-fits-all solutions. Even our interview participants, who are visualization researchers, are not immune to these issues.

We need to investigate whether a similar tendency exists in the application of other machine learning techniques for visual analytics. For instance, although we cannot always guarantee the faithfulness of LLMs (e.g., due to hallucinations [76]), they are commonly used in many applications without proper performance evaluation, justified by their perceived faithfulness in creating human-like responses [25]. Investigating whether LLMs are applied to appropriate tasks and properly parameterized is thus necessary. This is especially crucial since LLMs are not exclusively employed by experts but are increasingly used by non-experts.

Pursuing this direction is important given that advanced machine learning techniques like LLMs are becoming part of our everyday visual analytics [13]. As our research does for *t*-SNE and UMAP, such an examination will help develop solutions that support the proper use of these techniques and stimulate related discourse, contributing to enhancing the reliability of visual analytics.

9 Conclusion

We critically examine the misuse of *t*-SNE and UMAP in visual analytics. Through the literature review and the interview study, we verify the existence of the misuse and also reveal why such misuse occurs. We reveal that the misuses occur mainly because of the lack of discourse on the appropriate use of DR. Then, through the interview with DR experts, we discovered that existing attempts fail to motivate practitioners to cultivate their DR literacy. Based on these findings, we reluctantly suggest considering the automation of selecting DR projections in visual analytics as a potential solution to address misuse. Our study not only contributes to addressing the misuse of *t*-SNE and UMAP, but also encourages broader discussions on adopting a more critical perspective when applying machine learning techniques.

References

[1] Ehsan Amid and Manfred K. Warmuth. 2022. TriMap: Large-scale Dimensionality Reduction Using Triplets. arXiv:1910.00204 [cs.LG] https://arxiv.org/abs/1910.00204

[2] Sanjeev Arora, Wei Hu, and Pravesh K. Kothari. 2018. An Analysis of the t-SNE Algorithm for Data Visualization. In *31st Conference On Learning Theory*, Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet (Eds.), Vol. 75. 1455–1462. https://proceedings.mlr.press/v75/arora18a.html

[3] Daniel Atzberger, Tim Cech, Willy Scheibel, Jürgen Döllner, Michael Behrisch, and Tobias Schreck. 2025. A Large-Scale Sensitivity Analysis on Latent Embeddings and Dimensionality Reductions for Text Spatializations. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 305–315. doi:10.1109/TVCG.2024.3456308

[4] Daniel Atzberger, Tim Cech, Matthias Trapp, Rico Richter, Willy Scheibel, Jürgen Döllner, and Tobias Schreck. 2024. Large-Scale Evaluation of Topic Models and Dimensionality Reduction Methods for 2D Text Spatialization. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 902–912. doi:10.1109/TVCG.2023.3326569

[5] Bárbara C. Benato, Alexandre X. Falcão, and Alexandru C. Telea. 2024. Linking Data Separation, Visual Separation, Classifier Performance Using Multidimensional Projections. In *Computer Vision, Imaging and Computer Graphics Theory and Applications*. Springer Nature Switzerland, Cham, 229–255.

[6] Jürgen Bernard, Marco Hutter, Matthias Zeppelzauer, Michael Sedlmair, and Tamara Munzner. 2021. ProSeCo: Visual analysis of class separation measures and dataset characteristics. *Computers & Graphics* 96 (2021), 48–60. doi:10.1016/j.cag.2021.03.004

[7] Matthew Brehmer, Michael Sedlmair, Stephen Ingram, and Tamara Munzner. 2014. Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences. In *Proc. of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (Paris, France) *(BELIV '14)*. 1–8. doi:10.1145/2669557.2669559

[8] Dylan Cashman, Mark Keller, Hyeon Jeon, Bum Chul Kwon, and Qianwen Wang. 2025. A Critical Analysis of the Usage of Dimensionality Reduction in Four Domains. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–20. doi:10.1109/TVCG.2025.3567989

[9] Tara Chari and Lior Pachter. 2023. The specious art of single-cell genomics. *PLOS Computational Biology* 19, 8 (08 2023), 1–20. doi:10.1371/journal.pcbi.1011288

[10] Angelos Chatzimparmpas, Rafael M. Martins, and Andreas Kerren. 2020. t-viSNE: Interactive Assessment and Interpretation of t-SNE Projections. *IEEE Transactions on Visualization and Computer Graphics* 26, 8 (2020), 2696–2714. doi:10.1109/TVCG.2020.2986996

[11] A. Chatzimparmpas, F. V. Paulovich, and A. Kerren. 2023. HardVis: Visual Analytics to Handle Instance Hardness Using Undersampling and Oversampling Techniques. *Computer Graphics Forum* 42, 1 (2023), 135–154. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14726 doi:10.1111/cgf.14726

[12] Longfei Chen, Chen Cheng, He Wang, Xiyuan Wang, Yun Tian, Xuanwu Yue, Wong Kam-Kwai, Haipeng Zhang, Suting Hong, and Quan Li. 2024. FMLens: Towards Better Scaffolding the Process of Fund Manager Selection in Fund Investments. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–17. doi:10.1109/TVCG.2024.3394745

[13] Nan Chen, Yuge Zhang, Jiahang Xu, Kan Ren, and Yuqing Yang. 2025. VisEval: A Benchmark for Data Visualization in the Era of Large Language Models. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (2025), 1301–1311. doi:10.1109/TVCG.2024.3456320

[14] Furui Cheng, Mark S Keller, Huamin Qu, Nils Gehlenborg, and Qianwen Wang. 2023. Polyphony: an Interactive Transfer Learning Framework for Single-Cell Data Analysis. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 591–601. doi:10.1109/TVCG.2022.3209408

[15] Jaegul Choo, Hanseung Lee, Jaeyeon Kihm, and Haesun Park. 2010. iVisClassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *2010 IEEE Symposium on Visual Analytics Science and Technology*. 27–34. doi:10.1109/VAST.2010.5652443

[16] Andy Coenen and Adam Pearce. 2019. Understanding umap. *Google PAIR* (2019).

[17] Seoyoung Doh, Hyeon Jeon, Sungbok Shin, Ghulam Jilani Quadri, Nam Wook Kim, and Jinwook Seo. 2025. Understanding Bias in Perceiving Dimensionality Reduction Projections. arXiv:2507.20805 [cs.HC] https://arxiv.org/abs/2507.20805

[18] Alex Endert, Patrick Fiaux, and Chris North. 2012. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. 473–482. doi:10.1145/2207676.2207741

[19] Mateus Espadoto, Rafael M. Martins, Andreas Kerren, Nina S. T. Hirata, and Alexandru C. Telea. 2021. Toward a Quantitative Survey of Dimension Reduction Techniques. *IEEE Transactions on Visualization and Computer Graphics* 27, 3 (2021), 2153–2173. doi:10.1109/TVCG.2019.2944182

[20] Ronak Etemadpour, Robson Motta, Jose Gustavo de Souza Paiva, Rosane Minghim, Maria Cristina Ferreira de Oliveira, and Lars Linsen. 2015. Perception-Based Evaluation of Projection Methods for Multidimensional Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* 21, 1 (2015), 81–94. doi:10.1109/TVCG.2014.2330617

[21] Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. 2015. Efficient and robust automated machine learning. *Advances in neural information processing systems* 28 (2015).

[22] Takanori Fujiwara, Yun-Hsin Kuo, Anders Ynnerman, and Kwan-Liu Ma. 2023. Feature Learning for Nonlinear Dimensionality Reduction toward Maximal Extraction of Hidden Patterns. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*. 122–131. doi:10.1109/PacificVis56936.2023.00021

[23] Takanori Fujiwara, Xinhai Wei, Jian Zhao, and Kwan-Liu Ma. 2022. Interactive Dimensionality Reduction for Comparative Analysis. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 758–768. doi:10.1109/TVCG.2021.3114807

[24] Leo A. Goodman. 1961. Snowball Sampling. *The Annals of Mathematical Statistics* 32, 1 (1961), 148–170. http://www.jstor.org/stable/2237615

[25] Yi He, Ke Xu, Shixiong Cao, Yang Shi, Qing Chen, and Nan Cao. 2025. Leveraging Foundation Models for Crafting Narrative Visualization: A Survey. *IEEE*

Transactions on Visualization and Computer Graphics (2025), 1–20. doi:10.1109/TVCG.2025.3542504

[26] Geoffrey Hinton and Sam Roweis. 2002. Stochastic Neighbor Embedding. In *Proc. of the 15th International Conference on Neural Information Processing Systems*. MIT Press, Cambridge, MA, USA, 857–864.

[27] Fred Hohman, Kanit Wongsuphasawat, Mary Beth Kery, and Kayur Patel. 2020. Understanding and Visualizing Data Iteration in Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. 1–13. doi:10.1145/3313831.3376177

[28] Hyein Hong, Sangbong Yoo, Yejin Jin, and Yun Jang. 2023. How Can We Improve Data Quality for Machine Learning? A Visual Analytics System using Data and Process-driven Strategies. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*. 112–121. doi:10.1109/PacificVis56936.2023.00020

[29] Hyeon Jeon, Michaël Aupetit, Soohyun Lee, Hyung-Kwon Ko, Youngtaek Kim, and Jinwook Seo. 2022. Distortion-Aware Brushing for Interactive Cluster Analysis in Multidimensional Projections. doi:10.48550/ARXIV.2201.06379 (arXiv preprint).

[30] Hyeon Jeon, Hyung-Kwon Ko, Jaemin Jo, Youngtaek Kim, and Jinwook Seo. 2021. Measuring and Explaining the Inter-Cluster Reliability of Multidimensional Projections. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 551–561. doi:10.1109/TVCG.2021.3114833

[31] Hyeon Jeon, Hyung-Kwon Ko, Soohyun Lee, Jaemin Jo, and Jinwook Seo. 2022. Uniform Manifold Approximation with Two-phase Optimization. In *2022 IEEE Visualization and Visual Analytics (VIS)*. 80–84. doi:10.1109/VIS54862.2022.00025

[32] Hyeon Jeon, Kwon Ko, Soohyun Lee, Jake Hyun, Taehyun Yang, Gyehun Go, Jaemin Jo, and Jinwook Seo. 2025. UMATO: Bridging Local and Global Structures for Reliable Visual Analytics with Dimensionality Reduction. *IEEE Transactions on Visualization and Computer Graphics* (2025), 1–18. doi:10.1109/TVCG.2025.3602735

[33] Hyeon Jeon, Yun-Hsin Kuo, Michaël Aupetit, Kwan-Liu Ma, and Jinwook Seo. 2024. Classes are Not Clusters: Improving Label-Based Evaluation of Dimensionality Reduction. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 781–791. doi:10.1109/TVCG.2023.3327187

[34] Hyeon Jeon, Hyunwook Lee, Yun-Hsin Kuo, Taehyun Yang, Daniel Archambault, Sungahn Ko, Takanori Fujiwara, Kwan-Liu Ma, and Jinwook Seo. 2025. Unveiling High-dimensional Backstage: A Survey for Reliable Visual Analytics with Dimensionality Reduction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3706598.3713551

[35] Hyeon Jeon, Jeongin Park, Soohyun Lee, Dae Hyun Kim, Sungbok Shin, and Jinwook Seo. 2025. Dataset-Adaptive Dimensionality Reduction. arXiv:2507.11984 [cs.HC] https://arxiv.org/abs/2507.11984

[36] Hyeon Jeon, Ghulam Jilani Quadri, Hyunwook Lee, Paul Rosen, Danielle Albers Szafir, and Jinwook Seo. 2024. CLAMS: A Cluster Ambiguity Measure for Estimating Perceptual Variability in Visual Clustering. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 770–780. doi:10.1109/TVCG.2023.3327201

[37] Jaemin Jo, Jinwook Seo, and Jean-Daniel Fekete. 2020. PANENE: A Progressive Algorithm for Indexing and Querying Approximate k-Nearest Neighbors. *IEEE Transactions on Visualization and Computer Graphics* 26, 2 (2020), 1347–1360. doi:10.1109/TVCG.2018.2869149

[38] Minsuk Kahng, Pierre Y. Andrews, Aditya Kalro, and Duen Horng Chau. 2018. ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 88–97. doi:10.1109/TVCG.2017.2744718

[39] Minsuk Kahng, Nikhil Thorat, Duen Horng Chau, Fernanda B. Viégas, and Martin Wattenberg. 2019. GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 310–320. doi:10.1109/TVCG.2018.2864500

[40] Hyung-Kwon Ko, Jaemin Jo, and Jinwook Seo. 2020. Progressive Uniform Manifold Approximation and Projection. In *EuroVis 2020 - Short Papers*, Andreas Kerren, Christoph Garth, and G. Elisabeta Marai (Eds.). The Eurographics Association. doi:10.2312/evs.20201061

[41] Dmitry Kobak and George C. Linderman. 2021. Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nature Biotechnology* 39, 2 (01 Feb 2021), 156–157. doi:10.1038/s41587-020-00809-z

[42] Yun-Hsin Kuo, Takanori Fujiwara, Charles C.-K. Chou, Chun-houh Chen, and Kwan-Liu Ma. 2022. A Machine-learning-Aided Visual Analysis Workflow for Investigating Air Pollution Data. In *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*. 91–100. doi:10.1109/PacificVis53943.2022.00018

[43] Bum Chul Kwon, Hannah Kim, Emily Wall, Jaegul Choo, Haesun Park, and Alex Endert. 2017. AxiSketcher: Interactive Nonlinear Axis Mapping of Visualizations through User Drawings. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 221–230. doi:10.1109/TVCG.2016.2598446

[44] Jan Lause, Philipp Berens, and Dmitry Kobak. 2024. The art of seeing the elephant in the room: 2D embeddings of single-cell data do make sense. *bioRxiv* (2024). arXiv:https://www.biorxiv.org/content/early/2024/07/31/2024.03.26.586728.full.pdf doi:10.1101/2024.03.26.586728

[45] John A. Lee and Michel Verleysen. 2011. Shift-invariant similarities circumvent distance concentration in stochastic neighbor embedding and variants. *Procedia Computer Science* 4 (2011), 538–547. doi:10.1016/j.procs.2011.04.056

[46] Seongmin Lee, Benjamin Hoover, Hendrik Strobelt, Zijie J. Wang, ShengYun Peng, Austin Wright, Kevin Li, Haekyu Park, Haoyang Yang, and Duen Horng Polo Chau. 2024. Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion. In *2024 IEEE Visualization and Visual Analytics (VIS)*. 96–100. doi:10.1109/VIS55277.2024.00027

[47] Yiran Li, Junpeng Wang, Xin Dai, Liang Wang, Chin-Chia Michael Yeh, Yan Zheng, Wei Zhang, and Kwan-Liu Ma. 2023. How Does Attention Work in Vision Transformers? A Visual Analytics Attempt. *IEEE Transactions on Visualization and Computer Graphics* 29, 6 (2023), 2888–2900. doi:10.1109/TVCG.2023.3261935

[48] Leland McInnes, John Healy, and James Melville. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 [stat.ML] doi:10.48550/arXiv.1802.03426

[49] Linhao Meng, Stef van den Elzen, Nicola Pezzotti, and Anna Vilanova. 2024. Class-Constrained t-SNE: Combining Data Features and Class Probabilities. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 164–174. doi:10.1109/TVCG.2023.3326600

[50] Jacob Miller, Vahan Huroyan, Raymundo Navarrete, Md Iqbal Hossain, and Stephen Kobourov. 2024. ENS-t-SNE: Embedding Neighborhoods Simultaneously t-SNE. In *2024 IEEE 17th Pacific Visualization Conference (PacificVis)*. 222–231. doi:10.1109/PacificVis60374.2024.00032

[51] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. 2020. Topological Autoencoders. In *Proceedings of the 37th International Conference on Machine Learning*, Vol. 119. 7045–7054. https://proceedings.mlr.press/v119/moor20a.html

[52] Cristina Morariu, Adrien Bibal, Rene Cutura, Benoît Frénay, and Michael Sedlmair. 2023. Predicting User Preferences of Dimensionality Reduction Embedding Quality. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 745–755. doi:10.1109/TVCG.2022.3209449

[53] Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. 2019. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 438–448. doi:10.1109/TVCG.2018.2865240

[54] Ashwin Narayan, Bonnie Berger, and Hyunghoon Cho. 2021. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nature biotechnology* 39, 6 (2021), 765–774. doi:10.1038/s41587-020-00801-7

[55] Corey Nolet, Avantika Lal, Rajesh Ilango, Taurean Dyer, Rajiv Movva, John Zedlewski, and Johnny Israeli. 2022. Accelerating single-cell genomic analysis with GPUs. *bioRxiv* (2022). arXiv:https://www.biorxiv.org/content/early/2022/05/28/2022.05.26.493607.full.pdf doi:10.1101/2022.05.26.493607

[56] Luis Gustavo Nonato and Michaël Aupetit. 2019. Multidimensional Projection for Visual Analytics: Linking Techniques with Distortions, Tasks, and Layout Enrichment. *IEEE Transactions on Visualization and Computer Graphics* 25, 8 (2019), 2650–2673. doi:10.1109/TVCG.2018.2846735

[57] Karl F.R.S. Pearson. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2, 11 (1901), 559–572. doi:10.1080/14786440109462720

[58] Nicola Pezzotti, Julian Thijssen, Alexander Mordvintsev, Thomas Höllt, Baldur Van Lew, Boudewijn P.F. Lelieveldt, Elmar Eisemann, and Anna Vilanova. 2020. GPGPU Linear Complexity t-SNE Optimization. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 1172–1181. doi:10.1109/TVCG.2019.2934307

[59] Ilya Ploshchik, Angelos Chatzimparmpas, and Andreas Kerren. 2023. MetaStackVis: Visually-Assisted Performance Evaluation of Metamodels. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*. 207–211. doi:10.1109/PacificVis56936.2023.00030

[60] Tim Sainburg, Leland McInnes, and Timothy Q. Gentner. 2021. Parametric UMAP Embeddings for Representation and Semisupervised Learning. *Neural Computation* 33, 11 (10 2021), 2881–2907. arXiv:https://direct.mit.edu/neco/article-pdf/33/11/2881/1966656/neco_a_01434.pdf doi:10.1162/neco_a_01434

[61] Michael Sedlmair, Matt Brehmer, Stephen Ingram, and Tamara Munzner. 2012. Dimensionality reduction in the wild: Gaps and guidance. *Dept. Comput. Sci., Univ. British Columbia* (2012), 1–10.

[62] Parikshit Solunke, Vitoria Guardieiro, João Rulff, Peter Xenopoulos, Gromit Yeuk-Yin Chan, Brian Barr, Luis Gustavo Nonato, and Claudio Silva. 2024. Mountaineer: Topology-Driven Visual Analytics for Comparing Local Explanations. *IEEE Transactions on Visualization and Computer Graphics* 30, 12 (2024), 7763–7775. doi:10.1109/TVCG.2024.3418653

[63] C. O. S. Sorzano, J. Vargas, and A. Pascual Montano. 2014. A survey of dimensionality reduction techniques. arXiv:1403.2877 [stat.ML]

[64] Julian Stahnke, Marian Dörk, Boris Müller, and Andreas Thom. 2016. Probing Projections: Interaction Techniques for Interpreting Arrangements and Errors of Dimensionality Reductions. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 629–638. doi:10.1109/TVCG.2015.2467717

[65] Mark van de Ruit, Markus Billeter, and Elmar Eisemann. 2022. An Efficient Dual-Hierarchy t-SNE Minimization. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 614–622. doi:10.1109/TVCG.2021.3114817

[66] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 86 (2008), 2579–2605.

[67] Laurens Van Der Maaten, Eric O Postma, H Jaap van den Herik, et al. 2009. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research* 10, 66-71 (2009), 13.

[68] Jarkko Venna, Jaakko Peltonen, Kristian Nybo, Helena Aidos, and Samuel Kaski. 2010. Information Retrieval Perspective to Nonlinear Dimensionality Reduction for Data Visualization. *Journal of Machine Learning Research* 11, 13 (2010), 451–490. doi:10.5555/1756006.1756019

[69] Elio Ventocilla and Maria Riveiro. [n. d.]. A comparative user study of visualization techniques for cluster analysis of multidimensional data sets. *Information Visualization* 19, 4 ([n. d.]), 318–338. doi:10.1177/1473871620922166

[70] Qianwen Wang, Sehi L'Yi, and Nils Gehlenborg. 2023. DRAVA: Aligning Human Concepts with Machine Learning Latent Dimensions for the Visual Exploration of Small Multiples. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Article 833, 15 pages. doi:10.1145/3544548.3581127

[71] Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. 2021. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization. *Journal of Machine Learning Research* 22, 201 (2021), 1–73. http://jmlr.org/papers/v22/20-1061.html

[72] Martin Wattenberg, Fernanda Viégas, and Ian Johnson. 2016. How to Use t-SNE Effectively. *Distill* (2016). doi:10.23915/distill.00002

[73] Jiazhi Xia, Linquan Huang, Weixing Lin, Xin Zhao, Jing Wu, Yang Chen, Ying Zhao, and Wei Chen. 2022. Interactive Visual Cluster Analysis by Contrastive Dimensionality Reduction. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–11. doi:10.1109/TVCG.2022.3209423

[74] Jiazhi Xia, Yuchen Zhang, Jie Song, Yang Chen, Yunhai Wang, and Shixia Liu. 2021. Revisiting Dimensionality Reduction Techniques for Visual Cluster Analysis: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. doi:10.1109/TVCG.2021.3114694

[75] Xiwei Xuan, Xiaoyu Zhang, Oh-Hyun Kwon, and Kwan-Liu Ma. 2022. VAC-CNN: A Visual Analytics System for Comparative Studies of Deep Convolutional Neural Networks. *IEEE Transactions on Visualization and Computer Graphics* 28, 6 (2022), 2326–2337. doi:10.1109/TVCG.2022.3165347

[76] Yue Zhang et al. 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. arXiv:2309.01219 [cs.CL] https://arxiv.org/abs/2309.01219

[77] Yang Zhang, Jisheng Liu, Chufan Lai, Yuan Zhou, and Siming Chen. 2024. Interpreting High-Dimensional Projections With Capacity. *IEEE Transactions on Visualization and Computer Graphics* 30, 9 (2024), 6038–6055. doi:10.1109/TVCG.2023.3324851

[78] Yuansheng Zhou and Tatyana O. Sharpee. 2022. Using Global t-SNE to Preserve Intercluster Data Structure. *Neural Computation* 34, 8 (07 2022), 1637–1651. arXiv:https://direct.mit.edu/neco/article-pdf/34/8/1637/2034896/neco_a_01504.pdf doi:10.1162/neco_a_01504