Hyeon Jeon, *Seoul National University* | Dissertation Proposal

Towards Reliable Machine Learning for Visual Analytics

Visual analytics is the science of analytical reasoning supported by interactive visual interfaces. As a scientific process, **visual analytics must be reliable**. It should enable analysts to make informed decisions and generate knowledge that accurately reflects the underlying data. **However, visual analytics often becomes unreliable because of the underlying machine learning (ML) algorithms.** The unreliability stems from a general overtrust towards ML. For example, while ML algorithms can yield incorrect results, practitioners frequently employ them without sufficient validation.

I aim to remedy overtrust in ML to enhance the reliability of ML applications in the visual analytics field (Figure A). I first found that practitioners use ML algorithms that mismatch with desired visual analytics tasks, which led me to design **ML algorithms** that better align with the tasks. I also identified and addressed erroneous assumptions that make **ML evaluations** or the **interaction** with visualizations unreliable. In the rest of my Ph.D. study, I will focus on **democratizing** these findings so that practitioners can more reliably use ML for visual analytics.

My research especially focuses on **dimensionality reduction** (DR) algorithms (Figure B). DR algorithms (e.g., t-SNE [13], UMAP [11]) take high-dimensional data as input and generate low-dimensional projections that retain the structural properties of the input data, such as cluster structures (Figure B, blue ellipses). I focused on DR for two reasons. First, as DR provides visual summaries of complicated data, it often serves as a core VA component to study data patterns [12]. Concentrating on DR thus amplifies the practical impact of my research. Second, visual analytics using DR can easily be unreliable. DR projects data from the vast high-dimensional space to the narrow low-dimensional space. Therefore, resulting projections are inherently distorted [4, 12]. For instance, two or more clusters in the original data may deceptively appear as a single, undivided cluster [4, 6]. In this situation, analysts may misinterpret the cluster structure, leading to erroneous conclusions about the underlying data. For example, analysts may conclude that two semantically distinct classes should be merged although they are well separated in the original data.

My endeavor has led to contributions at prestigious venues in the field of visualization. Throughout the past three and a half years, I published **five first-authored papers in IEEE VIS and IEEE TVCG** [3, 4, 5, 6, 7] with one **best papers honorable mention award** [7]. I am striving to finalize the essential but still missing elements of my dissertation (see work-in-progress (wip) in Figure A).

Designing Reliable Algorithms

I tackled the assumption that practitioners regard UMAP [11] as a general-purpose DR algorithm. As UMAP emphasizes neighborhood structure while reducing dimensionality, it reliably supports analytic tasks related to local structure (e.g., identifying the nearest neighbors of a given point) but fails to do so for the tasks investigating global structure, such as comparing cluster density. However, UMAP is widely used not only for investigating local structures but also global structures in visual analytics.

I designed and implemented UMATO [5] as an alternative to UMAP that supports both local and global tasks reliably (Figure C). We separate UMAP's optimization into two distinct phases, each dedicated solely to local or global structures, allowing UMATO to accurately portray both. UMATO is not only published as a paper (**VIS 2022**) but also released as an open-source Python library. As of May 2024, UMATO received more than 100 stars on GitHub and has been downloaded more than 7K times.

Fostering Reliable Evaluations

I identified that practitioners often evaluate how well DR projections maintain cluster structure by assuming class labels to form ground truth clusters [9]. The evaluation is

A. Research Overview

I aim to tackle **prevalent assumptions** that harm the reliability of ML applications in visual analytics. This is done by **designing** and **democratizing** techniques for reliable ML applications.

Reliable Algorithms
- UMATO (**VIS 2022**)

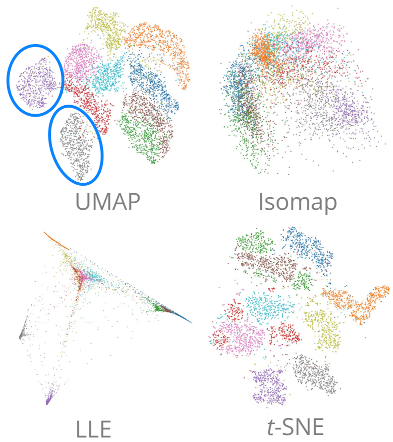
Reliable Evaluations
- Stead. & Cohev. (**VIS 2021**)
- Classes are not Clusters (**VIS 2023**)

Reliable Interactions
- CLAMS (**VIS 2023**)
  Best papers honorable mention (top 5%)
- Distortion-aware brushing (**wip**)
  under review (**TVCG**)

Democratizing Findings
- ZADU (**VIS 2023**)
- Structural Complexity (**wip**)
  under review (**VIS 2024**)
- Why not be reliable? (**wip**)

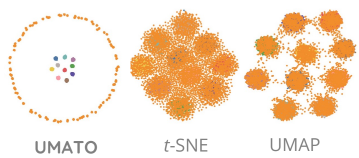
B. Diverse DR Projections of the MNIST dataset

Using DR projections, users can **visually grasp the overall distribution** of the original high-dimensional data. For example, they can identify **clusters** and examine their relationships.



C. Comparison btw UMATO and other DR Techniques

We project high-dimensional data comprised of **ten small 100D hyperspheres** and **a large 100-dim hypersphere enclosing the others (orange)** using DR techniques.



Only **UMATO** succeeds in **reliably representing the relationship between the hyperspheres**.

commonly done by quantifying how well the projections separate classes using clustering metrics, e.g., Silhouette. However, classes may not be clearly separated in the original high-dimensional data (Figure D), casting doubts about the reliability of DR evaluations using them as ground truth clusters (Figure E) [2, 9]. Since unreliable evaluation can result in an improper selection of a DR algorithm suited to the task or dataset, it's vital to address this unreliability.

Initially, I aimed to **exclude this assumption from DR evaluation by proposing to assess projections without using class labels**. I designed two DR quality metrics called Steadiness & Cohesiveness (S&C) for the purpose [4]. S&C leverage clustering algorithms (e.g., HDBSCAN) to detail cluster structure. Therefore, S&C capture distortions in cluster structure more accurately than not only ground-truth-based metrics but also other quality metrics for DR (e.g., Trustworthiness & Continuity [14]). S&C was presented in **VIS 2021**, where the corresponding paper was also published in **TVCG**.

Moreover, **I directly tackled the assumption by claiming that Classes are not Clusters** [6]. I designed Label-Trustworthiness and Label-Continuity (L-T&C), two DR quality metrics that employ class labels more reliably. Instead of assuming the classes as well-separated ground truth clusters, L-T&C quantifies their separability in the original data and assesses how well the separability is preserved in projections. L-T&C has proved effective in accurately quantifying distortions in DR projections, while the previous approach failed to detect evident distortions. The contribution of this work is acknowledged by a paper presented at **VIS 2023** and published in **TVCG**.

Making Interactions More Reliable

Visual analytics should anyway interact with users. In terms of DR, analysts first perceive the patterns prevalent in the scatterplots representing DR projections. Then, they can mark, annotate, or label interesting patterns using mouse or touch interaction. They can also directly control underlying DR algorithms, e.g., by changing distance metrics. In this context, I addressed two invalid assumptions that potentially make interactions with DR projections less reliable.

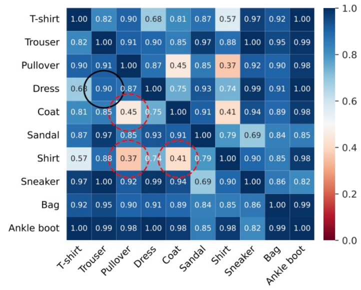
Assumption on the accuracy of DR projections. While designing visual analytics incorporating DR projections, practitioners often assume that the depicted structures accurately reflect the high-dimensional reality. This leads the system to guide users to "brush" the clusters they see using a rectangular box or lasso tool (Figure F, top). However, as mentioned earlier, clusters in DR projections may inaccurately reflect the original data distribution due to distortions. Therefore, clusters extracted by conventional brushing techniques may not stay as clusters in the original high-dimensional space.

I wanted to inform practitioners of the invalidity of the assumption and to build a brushing technique that works robustly against distortions. I thus designed Distortion-aware brushing [1], a novel brushing technique that complements DR projections to achieve reliable data analysis. Distortion-aware brushing accurately extracts high-dimensional clusters by resolving distortions around the brushed points by temporarily relocating points (Figure F, bottom), where our evaluation proved its superiority against previous brushing techniques. The work is currently under review in **TVCG**.

Assumption on the consistency of people. Another assumption I tackled is that the only factor that affects the reliability of visual analytics using DR projections is the accuracy of the projections. Based on such an assumption, practitioners optimize hyperparameter settings of DR algorithms while using the accuracy computed by DR quality metrics as a target function. However, in reality, unreliability also stems from humans. For example, visual analytics can be unreliable due to individual differences. When exposed to **ambiguous** visualizations, analysts can reach different conclusions [10, 15], making it uncertain which conclusion to accept. Of course, DR projections are not free from being ambiguous. The ways of perceiving clusters (i.e., conducting visual clustering) on DR projections can differ due to individual differences and unclear cluster boundaries (Figure G).

D. Pairwise Separability of Class Labels

The below heatmap depicts the **class-pairwise separability** of the Fashion-MNIST dataset computed using Distance Consistency [22]



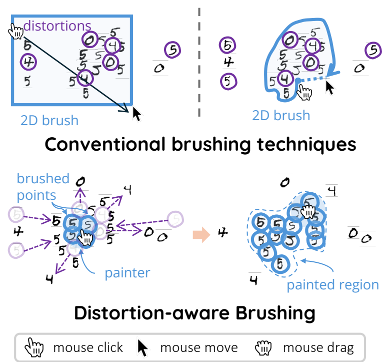
We can identify that ***Coat*, *Shirt*, *and Pullover*** classes have low pairwise separability, casting doubt on the reliability of using class labels as ground truth clusters.

E. A Pitfall of using Classes as Ground Truth Clusters

If **class labels** form well-separated **ground truth clusters**, we can easily distinguish good and bad DR projections by examining how well the classes are separated in the projections

If **class labels** are **not well separated**, using pairwise separability of classes in projections as a proxy for DR quality **becomes unreliable**.

F. Distortion-aware Brushing Illustration



Unlike conventional brushing techniques, **Distortion-aware brushing *relocates points*** *(purple arrows)* to locally resolve distortions, helping users to **identify clusters from DR projections more reliably**.

To address this assumption, I designed CLAMS [7], a visual quality metric that evaluates how ambiguous the cluster pattern of a given DR projection is. Trained on human perception data gathered with 34 participants [8], CLAMS accurately and scalably computes the ambiguity. By integrating CLAMS into the target function of DR optimization, I was able to produce less ambiguous DR projections with negligible accuracy loss (Figure G). The relevant paper was jointly published in the proceedings of **VIS 2023** and **TVCG**, winning a **Best Papers Honorable Mention Award**.

Democratizing the Findings: Towards Reliable ML for Everyone

Technology truly gains value when it is used by people. Although we now have diverse techniques that help the reliable use of ML for visualizations, they should be used by visualization researchers and designers. Accordingly, I strived to support more people in using ML for visual analytics in a more reliable manner. For example, I built ZADU, an open-source Python library providing DR quality metrics (Figure H) that achieved more than 6K downloads up to May 2024. The contribution of ZADU in making DR evaluation more accessible is acknowledged as a publication in **VIS 2023**. I also attempted to make DR hyperparameter optimization more efficient so that researchers could be more motivated to perform optimizations. This is done by measuring the structural complexity of high-dimensional datasets and putting less computational effort into simple datasets (under review in **VIS 2024**).

I plan to dedicate the rest of my Ph.D. study to further democratizing my findings. This will start by conducting an extensive literature survey to identify how researchers dealt with unreliability while using DR for visual analytics. Based on the survey, I will develop a self-checklist to help researchers identify threats that could compromise the reliability of their research. I will then interview visualization researchers to confront problems that obstruct researchers from being more reliable in using DR. Finally, I will design and develop tangible solutions that address these problems. For example, I would like to develop sanity-check software that can automatically spot potential reliability concerns in paper drafts.

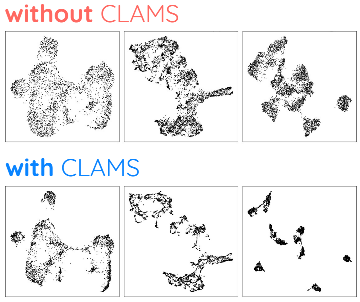
My ongoing work, which I plan to submit to ACM CHI 2025, is currently in the first stage of this plan, which is a literature survey. I expect the survey will construct a theoretical background that can ground future visual analytics research using not only DR but also other ML algorithms.

Future Research After Graduation

My passion for achieving reliable ML extends beyond DR and visual analytics. In my dissertation, I concentrate on one type of ML algorithm, dimensionality reduction (DR), to identify problems within the entire DR application pipeline. By doing so, I aim to build a solid foundation for improving the reliability of any ML algorithm. As a senior researcher, my future goal is to increase the reliability of a wider range of ML algorithms. For example, I would like to design reliable and efficient metrics that can detect hallucinations or deficiencies of multimodal foundation ML models (e.g., large language models).

G. Making DR less ambiguous with CLAMS

By combining **CLAMS into a target function** of DR optimization, we can produce DR projections with **less ambiguity** with a negligible loss of accuracy (DR quality metric scores).



H. ZADU Code Snippet

With ZADU, practitioners can assess the reliability of DR projections **by writing a few lines of code**.

```python
from zadu import zadu

hd, ld = load_datasets()
spec = [{
  "id" : "tnc",
  "params": { "k": 20 },
}]

zadu_obj = zadu.ZADU(spec, hd)
scores = zadu_obj.measure(ld)
print("T&C:", scores[0])
```

ZADU is released as an **open-source Python library** downloadable via pip. You can find the source code at:

github.com/hj-n/zadu

References

[1] **Hyeon Jeon**, Michaël Aupetit, Soohyun Lee, Hyung-Kwon Ko, Youngtaek Kim, and Jinwook Seo. Distortion-aware brushing for interactive cluster analysis in multidimensional projections, 2022. (arXiv preprint).

[2] **Hyeon Jeon**, Michael Aupetit, DongHwa Shin, Aeri Cho, Seokhyeon Park, and Jinwook Seo. Sanity check for external clustering validation benchmarks using internal validation measures, 2022. (arXiv preprint).

[3] **Hyeon Jeon**, Aeri Cho, Jinhwa Jang, Soohyun Lee, Jake Hyun, Hyung-Kwon Ko, Jaemin Jo, and Jinwook Seo. Zadu: A python library for evaluating the reliability of dimensionality reduction embeddings. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pages 196–200, 2023.

[4] **Hyeon Jeon**, Hyung-Kwon Ko, Jaemin Jo, Youngtaek Kim, and Jinwook Seo. Measuring and explaining the inter-cluster reliability of multidimensional projections. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):551–561, 2021.

[5] **Hyeon Jeon**, Hyung-Kwon Ko, Soohyun Lee, Jaemin Jo, and Jinwook Seo. Uniform manifold approximation with two-phase optimization. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pages 80–84, 2022.

[6] **Hyeon Jeon**, Yun-Hsin Kuo, Michaël Aupetit, Kwan-Liu Ma, and Jinwook Seo. Classes are not clusters: Improving label-based evaluation of dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):781–791, 2024.

[7] **Hyeon Jeon***, Ghulam Jilani Quadri*, Hyunwook Lee, Paul Rosen, Danielle Albers Szafir, and Jinwook Seo. Clams: A cluster ambiguity measure for estimating perceptual variability in visual clustering. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):770–780, 2024. (*: equal contribution).

[8] Mostafa M. Abbas, Michaël Aupetit, Michael Sedlmair, and Halima Bensmail. Clustme: A visual quality measure for ranking monochrome scatterplots based on cluster patterns. *Computer Graphics Forum*, 38(3):225–236, 2019.

[9] Michaël Aupetit. Sanity check for class-coloring-based evaluation of dimension reduction techniques. In *Proc. of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, page 134–141, 2014.

[10] Cindy Xiong Bearfield, Lisanne van Weelden, Adam Waytz, and Steven Franconeri. Same data, diverging perspectives: The power of visualizations to elicit competing interpretations. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–11, 2024.

[11] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.

[12] Luis Gustavo Nonato and Michaël Aupetit. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE Transactions on Visualization and Computer Graphics*, 25(8):2650–2673, 2019.

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

[14] Jarkko Venna and Samuel Kaski. Local multidimensional scaling. *Neural Networks*, 19(6):889–899, 2006.

[15] Cindy Xiong, Lisanne Van Weelden, and Steven Franconeri. The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3051–3062, 2020.