Immersive 3D Manifolds from Deep Embeddings: Evaluation of Image Classifications in VR

Meaghan Boykin[1], **Guilherme Melo dos Santos**[2], **Carlos Gurjão de Godoy**[2],
Regina C. Coelho[2], **Daniela Ushizima**[3,4]

[1]Comp. Sci. Department – Sonoma State University, Rohnert Park, CA – USA

[2]Comp. Sci. Department – Univ. Federal de São Paulo, S.J. dos Campos, SP – Brazil

[3]Comp. Division – Lawrence Berkeley National Laboratory, Berkeley, CA – USA

[4]Bakar Institute – University of California San Francisco, San Francisco, CA – USA

`mboykin@sonoma.edu, {guilherme.melo,gurjao.godoy,rccoelho}@unifesp.br,`

`dushizima@lbl.gov`

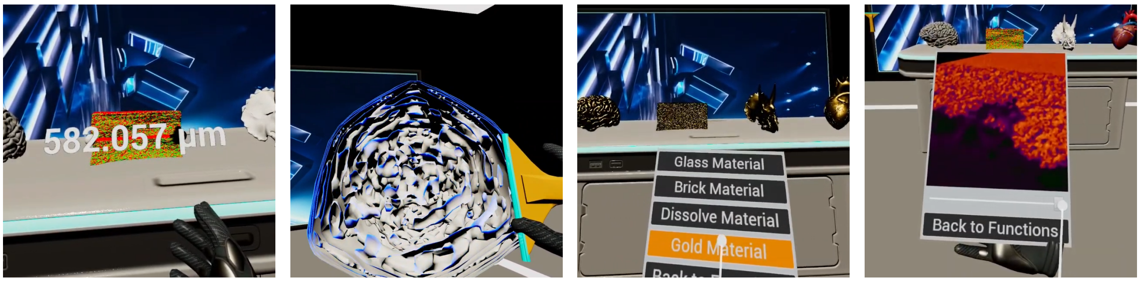
Abstract. *We show that our Meta Quest 3/3S application provides an immersive environment for the exploration of 3D manifolds generated from deep learning embeddings. Built with Unreal Engine 5.5.4, our tool enables the evaluation and sorting of biomedical specimens classifications by projecting high-dimensional semantic features into a navigable 3D space. Within the virtual environment, users can validate automated results by interactively categorizing images into defined spatial clusters, with the ability to export refined datasets via CSV. This systematic workflow improves quality control for large-scale imaging, offering researchers an intuitive interface to rapidly sort complex biological datasets.*

1. Introduction

Modern automated systems require Human-in-the-loop (HITL) integration to enhance performance and transparency [Barboza et al. 2025, Wu et al. 2022, van der Stappen and Funk 2021]. However, traditional 2D interfaces limit the integration of human intuition in complex computer vision tasks. ASCRIBE-VR (Autonomous Solutions for Computational Research with Immersive Browsing & Exploration in Virtual Reality) addresses this by providing a 3D immersive environment built with Unreal Engine 5.5.4 for the Meta Quest 3/3S. Expanding from its original mesh analysis tools (Fig. 1), this paper explains how the platform now enables the evaluation of biomedical specimens classifications through deep learning embeddings. By projecting high-dimensional semantic features into navigable 3D manifolds, users can validate automated results by manually sorting images into spatial clusters. This systematic workflow provides a high-throughput solution for quality control and CSV-based data refinement, significantly accelerating the validation of large-scale biological datasets [Ushizima et al. 2025].

2. Materials & Methods

The new HITL interface was developed using Unreal Engine's Blueprints Visual Scripting and tested on Meta Quest 3/3S. We utilized the Real Time Import/Export Plugin to engineer menu for runtime CSV and image loading, including other classes as it follows:



(a) Measure. (b) Slicing. (c) Materials. (d) Stacks.

Figure 1. Our application, ASCRIBE-VR, and its 3D visualization tools for specimen exploration: (a) ruler, (b) 3D slicer, (c) textures, (d) sample sections.

Image Actor: A specialized Blueprint class that displays images and metadata (Fig. 2). It includes a grab component and a widget component that toggles metadata visibility during interaction. Images use the `Find Look at Rotation` function to face the user automatically. **Image Manager:** An invisible actor that parses CSV files (containing filenames, 3D positions, and classifications) and spawns actors accordingly. **Cluster Area:** A 3D collision box that reclassifies any image actor entering its volume via `On Component Begin Overlap` events. When an image exits, its classification reverts to the original source value.

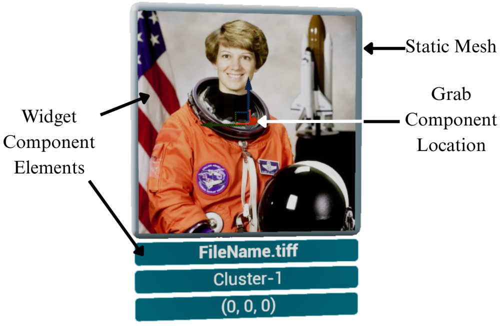


Figure 2. Image actor Blueprint showing the static mesh and metadata widget.

3. Data Processing Pipeline

To facilitate immersive exploration, raw images are transformed into a structured format containing semantic embeddings as spatial coordinates. This pipeline consists of feature extraction, dimensionality reduction via manifold learning, and density-based clustering.

3.1. Feature Extraction and Model Modularity

In the HITL app within ASCRIBE-VR, we used a deep convolutional neural network (CNN) to extract high-dimensional semantic features rather than to segment scientific images [Alegro et al. 2017, Ramalho et al. 2015, Ushizima et al. 2020], although such methods are used in creating polygonal meshes as shown in Fig. 1. The image embedding pipeline is implemented using PyTorch Lightning [Paszke et al. 2019], providing a modular framework that allows researchers to easily swap the feature extraction backbone based on the complexity of the dataset. Using `torchvision.models`, we can instantiate various architectures with specialized weights, such as `resnet18` [He et al. 2016] for efficiency or `resnet50` for higher feature resolution, as in the example below:

```
backbone = resnet50(weights=ResNet50_Weights.DEFAULT)
feature_extractor = torch.nn.Sequential(*list(backbone.children())[:-1])
```

By removing the final classification layer (fully connected layer), we treat the network as a fixed feature extractor. Images are resized to 224×224 pixels and passed through the backbone, resulting in a feature vector $f \in R^d$, where $d = 512$ for ResNet-18 or $d = 2048$ for ResNet-50.

3.2. Manifold Projection: Why UMAP?

To map these high-dimensional embeddings into the 3D navigable space of the virtual environment, we apply Uniform Manifold Approximation and Projection (UMAP) [McInnes et al. 2018]. While linear methods like Principal Component Analysis (PCA) are computationally faster, they rely on linear correlations and often fail to capture the complex, non-linear structures present in high-dimensional image manifolds. PCA tends to prioritize global variance, which can cause distinct clusters of similar images to overlap in 3D space. In contrast, UMAP excels at preserving both local and global topological structures. This ensures that visually similar images (e.g., specific biological traits, morphology) remain proximal in the virtual environment, which is critical for the "smart sorting" and quality control tasks within ASCRIBE-VR. The algorithm is configured to project the embeddings into three components ($n = 3$).

3.3. Automated Clustering and Coordinate Rescaling

Initial classifications are generated using DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [Ester et al. 1996]. Unlike k-means, DBSCAN does not require a pre-defined number of clusters, allowing it to adapt to the natural distribution of the embeddings. It also identifies outliers as "noise" (label -1), which helps researchers isolate low-quality or anomalous data samples during the VR review process. The resulting coordinates are rescaled from UMAP space to the VR coordinate system via linear transformation. For a coordinate p, the rescaled position v is calculated as:

$$v = V_{\min} + \frac{(p - P_{\min}) \cdot (V_{\max} - V_{\min})}{P_{\max} - P_{\min}} \tag{1}$$

where $[V_{\min}, V_{\max}]$ are boundaries defined for the Unreal Engine scene. The final output is a CSV file used by ASCRIBE-VR to spawn actors at semantically relevant 3D positions.

Table 1. Parameters for Feature Extraction, Projection, and Clustering

Stage	Parameter	Value
Feature Extraction	Backbone Model	ResNet-18
	Feature Dimensions	512
UMAP Projection	Neighbors (n)	15
	Minimum Distance	0.1
	Components	3
DBSCAN Clustering	Epsilon (ϵ)	0.65
	Min. Samples	5
VR Bounds (cm)	X Range	$[-459, 1405]$
	Y Range	$[-950, 815]$
	Z Range	$[70, 180]$

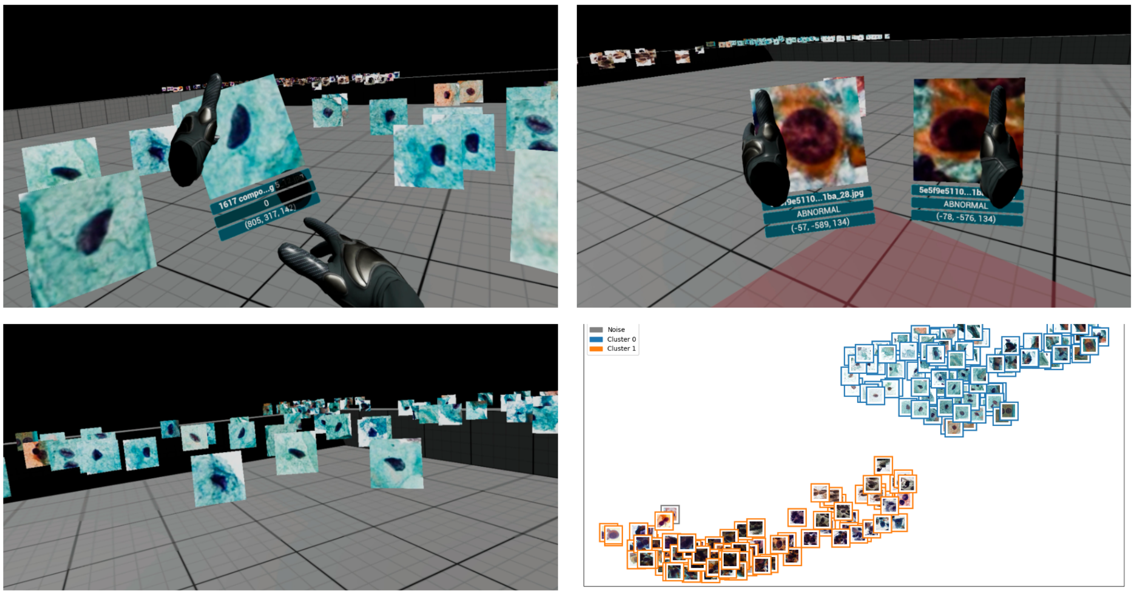


Figure 3. ASCRIBE-VR-HITL and respective 2D projection of embeddings manifold. Credits: Images from CRIC database [Rezende et al. 2021].

4. Results

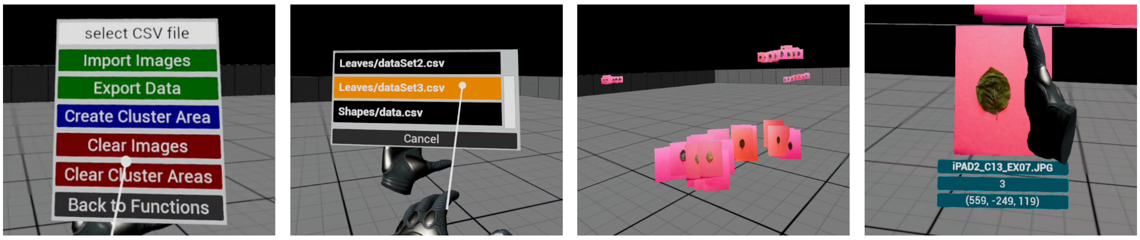
Figure 3 projects in 2D what will be visualized in VR, here the 3D manifold comes from cervical cells embeddings. After UMAP-reduction, each image gets a 3D coordinate, which is input to the ASCRIBE-VR HITL tool.

Researchers use the GUI (Fig. 4a-d) to load these embeddings and verify the semantic consistency of automated clinical clusters (Sec. 3.2–3.3). By deploying “Cluster Areas” (Fig. 4e-g), users rapidly correct AI misclassifications by physically re-assigning diagnostic labels to outlier specimens. This immersive approach facilitates the validation of complex biological datasets, exporting results via CSV for permanent record-keeping.

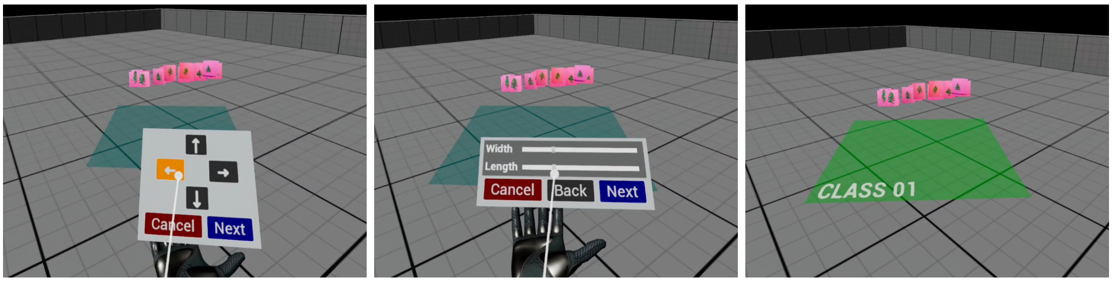
Fig. 5 illustrates the system performance, measuring how load times scale given image specimen count. To avoid hardware-specific “Texture Streaming Pool” errors at high resolutions, images were downsampled, enabling imports of up to 300 specimens. However, rendering overhead in the Meta Quest environment limits fluent interaction to approximately 250 active actors before lag-induced cybersickness occurs, establishing a practical threshold for immersive clinical review.

5. Discussion & Conclusion

This interface provides a robust framework for high-throughput image validation in biological research. By utilizing 3D proximity and spatial interaction, researchers can efficiently detect errors in automated segmentations that might be missed in 2D views. ASCRIBE-VR transforms data validation from a tedious labeling task into an intuitive spatial sorting workflow. Additional efforts have also enabled a Godot-based implementation with complementary utilities, called ASCRIBE-XR [Pandolfi et al. 2026]. Future iterations will focus on TIFF support and multi-image selection tools to further streamline clinical data quality control.



(a) Menu. **(b) Select.** **(c) 3D View.** **(d) Interact.**



(e) Placing. **(f) Sizing.** **(g) Clustered.**

Figure 4. The HITL engineered pipeline to (a-b) import new data, (c) view results from CNN/DBSCAN/UMAP clustering, (d) image metadata drops down upon grabbing, (e-g) HITL tools for data curation and class adjustments.

Acknowledgements

This work was supported by the US Department of Energy (DOE) Office of Science (SC) Advanced Scientific Computing Research (ASCR) and Basic Energy Sciences (BES) under Contract No. DE-AC02-05CH11231 to the Center for Advanced Mathematics for Energy Research Applications (CAMERA) program. It also included support from Autonomous Solutions for Computational Research with Immersive Browse & Exploration (ASCRIBE), Restor-C and Twins projects, which are supported by the SC of the U.S. Department of Energy under Contract No. DE-FOA-0003545.

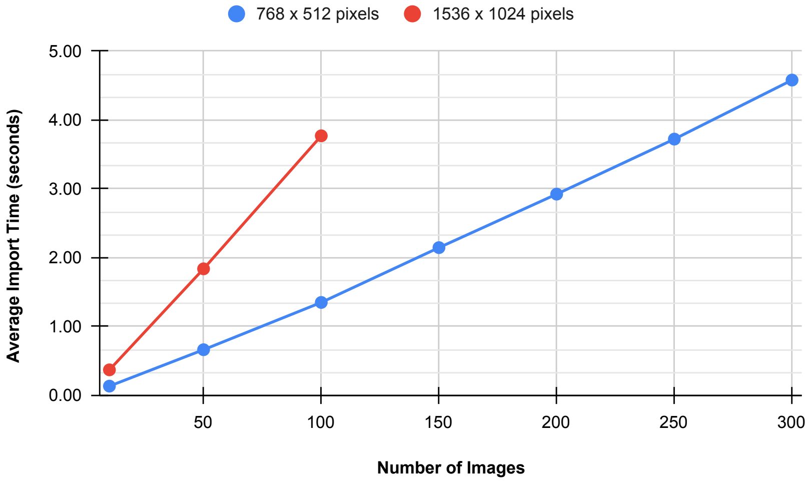


Figure 5. Average time taken to import images based on count and resolution.

References

Alegro, M., Theofilas, P., Nguy, A., Castruita, P. A., Seeley, W., Heinsen, H., Ushizima, D. M., and Grinberg, L. T. (2017). Automating cell detection and classification in human brain fluorescent microscopy images using dictionary learning and sparse coding. *Journal of neuroscience methods*, 282:20–33.

Barboza, L., Silva, F., Oliveira, H., Oliveira, J., and Brasil, A. (2025). Supporting stroke rehabilitation through gamified daily activity simulations. In *XXV SBCAS*. SBC.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, pages 226–231.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *Journal of Open Source Software*, 3(29):861.

Pandolfi, R. J., Todd, J., Donatelli, J., and Ushizima, D. (2026). Ascribe-xr: Extended reality for visualization of scientific images. In *Proc. 11th Int. Wrkshp on Data Analysis and Reduction for Big Scientific Data (DRBSD-11) at Super Computing (SC26).*

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 32.

Ramalho, G. L., Ferreira, D. S., Bianchi, A. G., Carneiro, C. M., Medeiros, F. N., and Ushizima, D. M. (2015). Cell reconstruction under voronoi and enclosing ellipses from 3d microscopy. In *IEEE International Symposium on Biomedical Imaging (ISBI).*

Rezende, M. T., Silva, R., Bernardo, F. d. O., Tobias, A. H. G., Oliveira, P. H. C., Machado, T. M., Costa, C. S., Medeiros, F. N. S., Ushizima, D. M., Carneiro, C. M., and Bianchi, A. G. C. (2021). Cric searchable image database as a public platform for conventional pap smear cytology data. *Nature Scientific Data*, 8(1):151.

Ushizima, D., dos Santos, G. M., Sordo, Z., Pandolfi, R., and Donatelli, J. (2025). Ascribe new dimensions to scientific data visualization and interactive exploration with VR. DOI:10.48550/arXiv.2504.13448.

Ushizima, D., Xu, K., and Monteiro, P. J. (2020). Materials data science for microstructural characterization of archaeological concrete. *MRS Advances*, 5(7):305–318.

van der Stappen, A. and Funk, M. (2021). Towards guidelines for designing human-in-the-loop machine training interfaces. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*, IUI '21, page 514–519, New York, NY, USA. Association for Computing Machinery.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., and He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135:364–381.