© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. This is the accepted version of: Rida Saghir, Saleem Ahmad, László Kopácsi, Thiago S. Gouvêa, Daniel Sonntag: User-Centric Active Learning Through Immersive Visualization. 2026 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2026. DOI: 10.1109/VRW70859.2026.00381

User-Centric Active Learning through Immersive Visualization

Rida Saghir[1,2*] Saleem Ahmad[2†] László Kopácsi[2‡] Thiago S. Gouvêa[2§] Daniel Sonntag[1,2¶]

[1]Applied Artificial Intelligence, University of Oldenburg, Oldenburg, Germany
[2]Interactive Machine Learning, German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

Abstract

Understanding data structure is critical for effective human involvement in machine learning, particularly in user-centric active learning, where annotation decisions by humans shape model behaviour. Learned embedding spaces capture this structure, but are typically explored through 2D interfaces. Virtual Reality (VR) offers an opportunity to present these embeddings immersively, supporting richer perception and interaction. We present a VR-based system for user-centric active learning that enables users to explore and annotate samples directly within an immersive embedding space. The system visualizes a continuously updating 3D embedding space derived from an audio classifier and augments it with common active learning cues, including uncertainty, density, diversity, and class coverage through visual encoding. User annotations trigger iterative retraining, allowing the model and visualization to co-evolve in response to human input.

Index Terms: Spatial Interaction, Embedding Visualization, Active Learning, Interactive Machine Learning

1 Introduction

VR is increasingly explored as a medium for data visualization as spatial immersion can make complex structures more perceivable and intuitive than traditional displays [5, 2]. Leveraging depth, embodied navigation, and natural manipulation, VR environments help users identify patterns, such as clusters and outliers, that are harder to discern in 2D. In machine learning, effective visualization is similarly critical for enabling human-in-the-loop interaction, yet immersive approaches remain rare. Limited VR-based systems include supporting model architecture inspection and manipulation [7], or interacting with learned embeddings [8], which are vector representations of data samples learned by the model and arranged spatially such that similar samples lie closer together. This limitation is especially apparent in active learning, a technique in which a model iteratively selects unlabeled samples based on some sampling queries for human annotation in order to improve performance under a limited labeling budget. While active learning explicitly relies on human input, users are often reduced to passive labelers, responding to model-centric queries without visibility into broader data distribution. Immersive visualization offers an opportunity to expose structure and promote sense-making, enabling more active and informed human participation. A growing body of work explores user-centric and mixed-initiative active learning interfaces in traditional 2D desktop settings, where humans are supported through embedding visualizations, model state cues, and interactive sample selection. In these systems, data samples are mapped into a low-dimensional embedding space, which can be interpreted similarly to a scatter plot; nearby points correspond to samples that the model considers similar, while clusters and separations reflect learned structure in the data. Prior work shows that when users can inspect such a data structure alongside model signals, human-driven selection can complement, or in some cases, outperform purely algorithmic sampling strategies in active learning [1, 3]. While 3D projections preserve more structure than 2D, their exploration on desktop displays remains constrained by 2D screen-based interaction. VR enables embodied, viewpoint-driven exploration of 3D embeddings supporting intuitive inspection of structure, relationships, and informative regions in the data. Although VR-based annotation environments have been explored outside of active learning contexts [4], active learning interfaces have largely remained confined to 2D settings. We therefore develop a VR-based interface for active learning that enables users to explore and annotate samples within an immersive 3D embedding space. The interface exposes data structure and augments it with model state cues using visual encoding to guide users towards informative samples and support focused exploration with dedicated filters.

*e-mail: rida.saghir@uni-oldenburg.de
†e-mail: saleem.ahmad@dfki.de
‡e-mail: laszlo.kopacsi@dfki.de
§e-mail: thiago.gouvea@dfki.de
¶e-mail: daniel.sonntag@dfki.de

2 System Design

The technical stack comprises a Python-based backend responsible for training the classifier and executing the active learning logic, coupled with a FastAPI server that communicates with the VR-based frontend via REST APIs. For the demo, we use the UrbanSound8K[1] dataset having 10 audio classes, initializing the system with a pretrained YAMNet[2] model and training a lightweight 256×3 dimensional neural classifier, enabling projection of learned representations into a three dimensional space for VR visualization. The model is first trained on a small set of labeled samples and produces both class predictions and embeddings. Users then inspect the embedding space in VR and annotate 10 samples, which triggers iterative retraining of the model. After each training round, the system updates the 3D embeddings along with the four commonly used active learning sampling cues: uncertainty, local density, diversity, and class coverage. Rather than using these cues to drive automatic sampling like in traditional model-centric active learning, we expose them visually to support human-guided selection of informative samples.

For each sample i with embedding $\mathbf{z}_i$ and predicted class probabilities $p_{i,c}$, *uncertainty* $u(i)$ captures model ambiguity and is measured using prediction entropy, $u(i) = -\sum_c p_{i,c} \log p_{i,c}$. *Local density rho*$(i)$ captures local sample concentration in the unlabeled embedding space and is computed as the inverse of the mean distance to the k nearest unlabeled neighbors, $\rho(i) = \left(\frac{1}{k}\sum_{j\in \mathrm{kNN}(i)} \|\mathbf{z}_i - \mathbf{z}_j\|\right)^{-1}$. Diversity $d(i)$ prioritizes samples that are distant from already labeled data, $\mathscr{L}$, $d(i) = \min_{l\in\mathscr{L}} \|\mathbf{z}_i - \mathbf{z}_l\|$. Class coverage $c(i)$ promotes balanced labeling by weighting samples whose predicted class $\hat{y}_i = \arg\max_c p_{i,c}$ is underrepresented in the labeled set, $c(i) = \left(N_{\mathscr{L}}(\hat{y}_i) + \varepsilon\right)^{-1}$.

[1] https://www.kaggle.com/datasets/chrisfilo/urbansound8kaudio
[2] https://www.tensorflow.org/hub/tutorials/yamnet

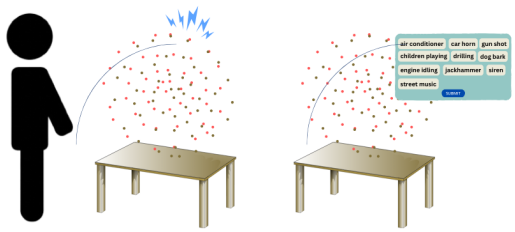

Figure 1: Conceptual illustration of the VR interaction flow, showing audio inspection (left) and the spatial annotation menu (right)

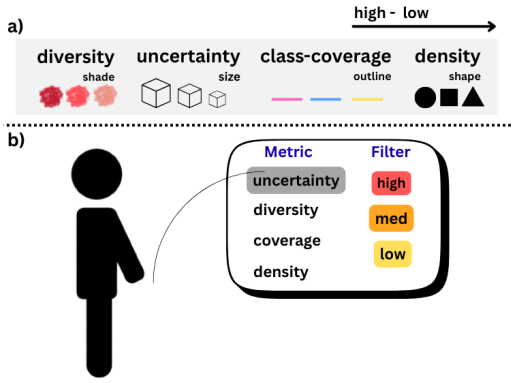

Figure 2: Visual encoding and filtering design. (a) Mapping of active learning cues to visual variables in the embedding space. (b) Illustration of embedding space filtering, allowing selection of high, medium, or low value ranges for each cue.

2.1 User Interface and Interaction Design

We use a tabletop-style VR view to render embeddings, drawing parallels to familiar desktop annotation workflows while avoiding the navigational complexity of open-world VR. Prior comparisons between immersive VR and desktop-based interfaces suggest that tabletop views support established interaction patterns, whereas immersive VR primarily enhances perceptual understanding in spatial tasks [2]. We render the learned representation of the model as 3D embeddings to preserve more of the model's latent structure than a 2D projection. Automatic rotation of the embedding space was avoided, as simultaneous movement of many data points introduced visual lag: in our implementation (Unity Editor on a laptop with an i7-11800H CPU, 16 GB RAM, and an NVIDIA RTX 3070 GPU), the system achieved approximately 25-35 FPS in static visualization and 15-25 FPS during embedding updates. Instead, users inspect the embedding through local viewpoint adjustments around the fixed tabletop reference and controller-based zoom. Interpretation of individual samples is facilitated through audio playback, which is triggered when users select a data point using a controller button. After inspection, annotation is performed using a spatial, ray-cast menu anchored to the selected point (Fig. 1). Prior evaluations show that spatial menus are faster and less physically demanding than body-referenced alternatives [6], and that menu shape has minimal impact on performance; accordingly, we adopt a simple rectangular layout. The menu applies yaw-only rotation to face the user, preserving readability without introducing tilt or spatial drift. To support focused exploration during annotation, we provide toggleable spatial filtering controls that expose common active learning cues, including uncertainty, diversity, density, and class coverage (Fig. 2). Embeddings are colored by the model's predicted class, following standard practice in embedding visualizations, while each active learning cue is discretized into three bins (low, medium, high) and represented using a distinct visual variable: size for uncertainty, color saturation for diversity, outline color for class coverage, and shape for density. Users activate bins via ray-casting to filter the embedding accordingly. Annotation proceeds in discrete rounds of 10 samples, with the current iteration number and model accuracy displayed and updated after each retraining round. After each round, the model retrains for one epoch on newly labeled samples (approximately 5-8 seconds in our setup), and the updated embeddings are linearly interpolated from their previous to new positions, producing gradual motion that allows users to visually track changes in the embedding space.

3 Conclusion and Future Work

This work represents an initial step toward immersive human-in-the-loop active learning, demonstrating how model embeddings and cues can be exposed in VR to support human-guided annotation. While the current system supports basic interaction and annotation to influence model behavior, future work will extend the interface with richer model feedback, progress indicators, and performance metrics, informed by prior literature on interactive and visual active learning. In parallel, we plan to conduct an initial pilot study to assess mental workload and interface understandability. Building on these findings, we will perform a comprehensive user study to evaluate annotation efficiency, including an A/B comparison with a functionally equivalent desktop-based interface, to better understand the strengths and limitations of immersive interaction.

Acknowledgments

This research is part of the Computational Sustainability & Technology project area[3], and has been supported by the Lower Saxony Ministry of Science and Culture (MWK) in zukunft.niedersachsen program, the Federal Ministry of Research, Technology and Space (BMFTR) under grant number 16IW23002 (No-IDLE), and the Endowed Chair of AAI at University of Oldenburg.

References

[1] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE transactions on visualization and computer graphics*, 24(1):298–308, 2017. 1

[2] A. Bueckle, K. Buehling, P. C. Shih, and K. Börner. 3d virtual reality vs. 2d desktop registration user interface comparison. *PloS one*, 16(10):e0258103, 2021. 1, 2

[3] M. Chegini, J. Bernard, J. Cui, F. Chegini, A. Sourin, K. Andrews, and T. Schreck. Interactive visual labelling versus active learning: an experimental comparison. *Frontiers of Information Technology & Electronic Engineering*, 21(4):524–535, 2020. 1

[4] A. Doula, T. Güdelhöfer, A. Matviienko, M. Mühlhäuser, and A. Sanchez Guinea. Immersive-labeler: Immersive annotation of large-scale 3d point clouds in virtual reality. In *ACM SIGGRAPH 2022 Posters*, pp. 1–2. 2022. 1

[5] D. Hayatpur, H. Xia, and D. Wigdor. Datahop: Spatial data exploration in virtual reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 818–828, 2020. 1

[6] I. Lediaeva and J. J. LaViola. Evaluation of body-referenced graphical menus in virtual environments. In *Graphics Interface*, vol. 2020, p. 6, 2020. 2

[7] Z. Lyu, J. Li, and B. Wang. Aiive: Interactive visualization and sonification of neural networks in virtual reality. In *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 251–255. IEEE, 2021. 1

[8] A. Prange and D. Sonntag. A demonstrator for interactive image clustering and fine-tuning neural networks in virtual reality. In *German Conference on Artificial Intelligence (Künstliche Intelligenz)*, pp. 194–203. Springer, 2021. 1

[3] https://cst.dfki.de/