PanoStyleVR: style-based similarity metrics for Web-based immersive panoramic style transfer

Muhammad Tukur*
College of Science & Engineering, Hamad Bin Khalifa University
Doha, Qatar
mtukur@hbku.edu.qa

Sara Jashari*
College of Science & Engineering, Hamad Bin Khalifa University
Doha, Qatar
sarajashari@infidev.mk

Dana Abou Hassanain
College of Science & Engineering, Hamad Bin Khalifa University
Doha, Qatar
daab51784@hbku.edu.qa

Fabio Bettio
Visual and Data-Intensive Computing, CRS4
Cagliari, Italy
fabio.bettio@crs4.it

Giovanni Pintore
Visual and Data-Intensive Computing, CRS4
Cagliari, Italy
giovanni.pintore@crs4.it

Enrico Gobbetti
Visual and Data-Intensive Computing, CRS4
Cagliari, Italy
enrico.gobbetti@crs4.it

Jens Schneider
College of Science & Engineering, Hamad Bin Khalifa University
Doha, Qatar
jeschneider@hbku.edu.qa

Marco Agus†
College of Science & Engineering, Hamad Bin Khalifa University
Doha, Qatar
magus@hbku.edu.qa

Figure 1: In PanoStyleVR a feature-based style similarity metric is used to organize a database of panoramic images of indoor environments. At run time, given a novel monocular 360° of a room, an immersive *WebXR*-based interface supports stereoscopic exploration (left), interactive selection (middle), and real-time application (right) of photorealistic style transfers from recommended images extracted from the database.

Abstract

We introduce *PanoStyleVR*, an immersive web-based framework for analyzing, ranking, and interactively applying style similarities within panoramic indoor scenes, enabling stereoscopic virtual exploration and photorealistic style adaptation. A key innovation of our system is a fully immersive *WebXR* interface, allowing users wearing head-mounted displays to navigate indoor environments in stereo and apply new styles in real time. Style suggestions are visualized through floating thumbnails rendered in the VR space; selecting a style triggers photorealistic transfer on the current room view and updates the immersive stereo representation. This interactive pipeline is powered by two integrated neural components: (1) a geometry-aware and shading-independent GAN-based framework for semantic style transfer on albedo-reflectance representations; and (2) a gated architecture that synthesizes omnidirectional stereoscopic views from a single 360° panorama for realistic depth-aware exploration. Our system enables cosine-similarity-based style ranking, t-SNE-driven dimensionality reduction, and GMM-based clustering over large-scale panoramic datasets. These components support an immersive recommendation mechanism that connects stylistic analysis with interactive editing. Experimental evaluations on the Structured3D dataset demonstrate strong alignment between

*Both authors contributed equally to this research.
†Corresponding author.

Web3D '25, Siena, Italy

ACM ISBN 979-8-4007-2038-3/2025/09
https://doi.org/10.1145/3746237.3746287

perceptual similarity and our proposed metric, and effective grouping of panoramas based on latent style representations.

CCS Concepts

• Computing methodologies → Graphics systems and interfaces; **Scene understanding**; **Neural networks**; **• Human-centered computing → Virtual reality**; **Web-based interaction**.

Keywords

Panoramic Indoor Scenes, Style Similarity Ranking, Style Transfer, Immersive recommendation system

ACM Reference Format:

Muhammad Tukur, Sara Jashari, Dana Abou Hassanain, Fabio Bettio, Giovanni Pintore, Enrico Gobbetti, Jens Schneider, and Marco Agus. 2025. PanoStyleVR: style-based similarity metrics for Web-based immersive panoramic style transfer . In *The 30th International Conference on 3D Web Technology (Web3D '25), September 9–10, 2025, Siena, Italy*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746237.3746287

1 Introduction

Virtual styling is a specialized form of virtual staging that focuses on updating or enhancing the interior design of a furnished space. Instead of focusing on adding furniture to an empty room, as with traditional virtual staging, it lets users reimagine the style of a lived-in or furnished space without accessing it or making physical changes. This application is significant for the real-estate market, where it has been shown to facilitate design, reduce cost, and accelerate communication with clients, especially when coupled with immersive presentations [Sharma 2024; Xiong et al. 2024]. This requires end-to-end solutions, encompassing capture, processing, editing, and virtual presentation. In this context, 360° images have become the de facto standard [Gobbetti et al. 2024]. Single 360° shots instantly capture the full scene from a single viewpoint, providing rich context for scene understanding [Pintore et al. 2024a; Yang et al. 2018] and, when consumed through Head-Mounted Displays (HMDs) they let users dynamically explore scenes with natural head motions, leading to good degrees of immersion and facilitating intuitive virtual reality (VR) interfaces [Matzen et al. 2017].

Understanding and modeling stylistic characteristics of indoor environments is a fundamental challenge in computer vision, immersive visualization, and virtual staging [Kim and Lee 2020; Naseer et al. 2019; Pintore et al. 2022]. Advances in neural style transfer (NST) [Jing et al. 2019], generative adversarial networks (GANs) [Goodfellow et al. 2014], and panoramic image processing [Tukur et al. 2023b; Zhi et al. 2022] have laid the groundwork for photorealistic panoramic-based content adaptation in indoor spaces. However, most existing methods treat style transfer as a static transformation task and lack a structured mechanism for exploring stylistic relationships across scenes or for deploying these capabilities in immersive, user-driven environments.

To address these limitations, we introduce *PanoStyleVR*, a novel web-based immersive framework for interactive exploration, ranking, and transfer of styles in panoramic indoor scenes. The framework comprises a front-end component for immersive exploration and dynamic style selection and application, and a backend component for intelligent style recommendation and application.

The system takes as input a single 360° shot of a real environment, transformed to a stereoscopic multiple-center-of-projection (MCOP) equirectangular image pair through PanoStereo [Pintore et al. 2024b], a recently introduced gated network architecture. At run-time, the pair is loaded in a lightweight *WebXR* viewer that responds to head rotations, offering both motion and stereo cues. During exploration, a small set of suggested candidate styles is displayed as floating thumbnails rendered within the 3D scene. Upon selecting one of these floating images, the corresponding style is transferred onto the environment, and a new set of suggested styles is proposed, enabling immediate immersive feedback. Style transfer is achieved through *PanoStyle* [Tukur et al. 2023b], a geometry-aware and shading-independent photorealistic style transfer model that operates on albedo reflectance representations to ensure consistency in indoor stylization. To support intelligent style recommendation and retrieval, *PanoStyleVR* builds on a backend style-similarity engine. Style codes are extracted using semantic-aware encoding grounded in the NYU_v2 object palette [Zheng et al. 2020], capturing semantic, geometric, and photometric attributes. We compute cosine similarity between style vectors to perform ranking and integrate dimensionality reduction (PCA, t-SNE) and clustering (DBSCAN, K-Means, GMM) to organize scenes with related stylistic characteristics. This enables an automated, explainable mechanism for immersive style suggestion and clustering.

Our work combines and extends recent work in panoramic image analysis and style transfer, introducing important contributions. In particular:

- We introduce an immersive *WebXR*-based interface that supports interactive selection and real-time application of photorealistic style transfers;
- We integrate PanoStyle [Tukur et al. 2023b] and PanoStereo [Pintore et al. 2024b] for geometry-aware stylization and stereo scene generation from a single panorama;
- We introduce a style ranking and clustering backend, leveraging NYU_v2-based codes, cosine similarity, t-SNE, and GMM for structured recommendation;

The rest of this paper is organized as follows. After briefly reviewing related work (Sec. 2), we summarize the design of our framework (Sec. 3). We then provide details on the two main components, i.e., the immersive *WebXR*-based interface (Sec. 4) and the style analysis and recommendation features (Sec. 5). We then provide a preliminary quantitative and qualitative evaluation of our prototypes (Sec. 6). The paper concludes with a summary of achievements and a view of the current and future work (Sec. 6).

2 Related work

Our work addresses the design of a feature-based style similarity metric for panoramic indoor scenes and the development of an immersive recommendation system for the interactive application of photorealistic styles. In the following, we briefly review the state-of-the-art in similarity metrics, style-based immersive recommendation systems, and immersive exploration, concentrating on the approaches closer to ours. For a broader overview, we refer the reader to recent surveys on indoor reconstruction and visualization [Pintore et al. 2020], panoramic indoor image processing and exploration [Gobbetti et al. 2024], and neural style transfer and associated evaluation metrics [Zhou et al. 2025].

Style similarity metrics for indoor scenes. Unlike artistic style transfer, which mostly focuses on emulating historical art movements or individual artist styles [Mao et al. 2017; Xu et al. 2025], our work focuses on interior design styles, which reflect both functional and aesthetic attributes and are defined by the cohesive arrangement of colors, materials, and layouts within a scene [Pintore et al. 2019]. Since we are interested in recommending style variations of the presented scene, we must compute style similarity metrics. Several methods have been proposed for measuring image similarity based on perceptual, structural, spectral, or textural features [Aguilera et al. 2016; Ding et al. 2020; Zujovic et al. 2013]. Traditional metrics include Root Mean Square Error (RMSE) [Sheikh and Bovik 2006] and Structural Similarity Index Measure (SSIM) [Wang et al. 2004], while recent deep learning-based approaches have incorporated semantic and texture-aware embeddings [Ding et al. 2020]. The Frechet Inception Distance (FID) [Obukhov and Krasnyanskiy 2020], widely used to assess the quality of generative models [Buzuti and Thomaz 2023], has also been adapted to evaluate stylistic coherence in artistic domains [Wright and Ommer 2022]. Specialized similarity metrics have also been extended to accommodate panoramic imagery and equirectangular distortions [Zhou et al. 2018]. More recently, Somepalli et al. [2024, 2025] introduced descriptor-based metrics for linking generated images (e.g., from Stable Diffusion) to the styles found in their training data. In contrast, our method operates on style embeddings generated via a geometry- and shading-aware GAN framework [Tukur et al. 2023b; Ye et al. 2025], tailored to indoor panoramic scenes. To the best of our knowledge, this work proposes the first framework to define and apply a style similarity metric for indoor style clustering and retrieval on equirectangular panoramic images.

Style-based immersive recommendation systems. Style similarity plays a critical role in the development of content-based recommendation and retrieval systems [Iqbal et al. 2019; Ko et al. 2022; Qazanfari et al. 2023]. While much prior work focuses on style in illustrations [Garces et al. 2014], artwork [Ruta et al. 2021; Shen et al. 2019], or shape-based analysis [Lun et al. 2015], including 3D furniture retrieval [Liu et al. 2015] and procedural room layout generation [Zhao et al. 2024], most methods lack immersive interaction capabilities and are not designed for photorealistic panoramic environments. Style-driven recommendation systems have been extensively explored in social media [Zhang and Yamasaki 2021], fashion [Deldjoo et al. 2023; Kachbal et al. 2024], outfit matching [De Divitiis et al. 2023], and cosmetic applications [Gulati et al. 2023; Sii and Chan 2025]. However, despite the growing interest in style-aware personalization, no previous work has proposed a system for indoor design that integrates immersive VR interaction, panoramic indoor style transfer, and a structured style similarity backend. To fill this gap, our work introduces the first immersive web-based recommendation system that supports interactive style selection and real-time panoramic style transfer, leveraging our novel indoor similarity metric.

Immersive exploration from monocular 360° input. Single-shot panoramas can replicate real-world environments but suffer from flat appearance [Matzen et al. 2017; Waidhofer et al. 2022]. To improve immersion, several methods have been proposed to support stereo-cues. By estimating depth, point cloud rendering [Huang et al. 2017], depth map-based meshes [Tukur et al. 2023a], and blended RGB-D data [Luo et al. 2018] can be used to render from displaced points, but lack details in disoccluded areas. Several view-synthesis methods based on deep learning have recently shown the ability to generate believable details, but their computation cost limits their direct application on embedded devices, forcing the use of remote rendering for HMDs [Pintore et al. 2023; Xu et al. 2021]. For these reasons, several authors have proposed precomputed approximations. Recent advances like layered depth images [Hedman and Kopf 2018], multi-plane panoramas (MPI) [Tucker and Snavely 2020], and adaptive sampling [Li and Khademi Kalantari 2020] have improved viewpoint flexibility. For stereo rendering, omnidirectional synthesis [Pintore et al. 2024b] generates equirectangular views optimized for VR, delivering strong stereo in central vision with some peripheral degradation—but applied only to original images. We extend this approach by integrating semantic analysis, enabling immersive stereo exploration of restyled panoramas.

3 Framework overview

The *PanoStyleVR* system comprises an immersive, *WebXR*-based editing interface for real-time style application and visualization (Sec. 4) that builds on a style encoding and similarity computation pipeline providing style clustering and adaptive recommendations (Sec. 5).

The back-office components of the framework create a database of the panoramas that provide stylistic examples. The panoramas are analyzed to extract style codes, which are used as a basis to compute similarity among styles and support clustering and similarity-based retrieval. At run time, given an input new panoramic indoor image, the system extracts its semantic and photometric features, encodes its style, and uses the code to retrieve visually the N top similar panoramas in the style database using a cosine-similarity metric. The panorama is also transformed into an explorable MCOP stereo pair to feed the *WebXR* viewer. The viewer exploits the MCOP image to provide stereoscopic immersive exploration of the environment, and the set of similar panoramas to create a visual menu of possible styling options. Users can explore candidate styles and apply them interactively; upon selection, the choice is communicated to the backend, which feeds the viewer with a new MCOP pair, with the style applied, and a new set of stylistic computed from the selected style.

The *WebXR* implementation of the viewer is detailed in Sec. 4, while the similarity metric at the basis of the recommendation system is detailed in Sec. 5.

4 Immersive styling application

The immersive editing system of *PanoStyleVR* is designed to enable interactive, real-time application of photorealistic styles within a VR environment. It is implemented using *WebXR* and *Three.js*, supporting a variety of head-mounted display devices.

Our system exploits the *PanoStereo* [Pintore et al. 2024b] method to automatically and rapidly convert a single panoramic image in equirectangular format into two aligned MCOP images encoded in equirectangular format. The method assumes that, during head rotations, each eye follows a circular arc. For each ambient to be immersively explored, which may be the original captured panorama or a stylized version, the stereoscopic effect is achieved by generating a per-eye MCOP image the pixel column for each longitude is generated from a different camera position, which corresponds to the position of the eye when it is looking straight in this direction. The two panoramas are transmitted to the headset each time the scene changes (i.e., at the change of the inspected room or the change of room styling), and rendering is accomplished using a *WebXR* viewer that sets the images as textures for two spheres positioned in the scene — one centered around the left eye and the other around the right eye. When XR rendering is enabled, the system retrieves at each frame the head's orientation from the headset's sensors, which is used to compute the view matrix for rendering the panoramas for each eye, thereby providing motion parallax and stereo perception.

When a user enters a virtual room rendered from a panoramic stereo pair, they are surrounded by a set of floating image thumbnails representing candidate styles. These suggestions are computed based on the similarity ranking of the current scene's style code against a pre-indexed database (see Sec. 5). Each thumbnail corresponds to a real indoor panorama whose style has been encoded and clustered in the backend. Upon selecting a style image by gazing or clicking, the system invokes the *Panostyle* backend to perform semantic-aware, geometry-consistent style transfer. The edited reflectance image is then recombined with the original shading signal to produce a photorealistic result. Subsequently, PanoStereo [Pintore et al. 2024b] is employed to generate a new stereoscopic panoramic pair from the stylized scene, which is streamed back to the *WebXR* client, allowing the user to seamlessly continue immersive navigation in the newly stylized room. Together with the updated image, the back-end also streams back to the viewer new thumbnails for suggested styles. This interaction loop can be repeated to explore multiple stylistic transformations in real time, enabling design iteration, personalization, or preference-based exploration.

The entire pipeline is modular and supports both offline precomputation and on-demand streaming. Preprocessing stages for reflectance and semantic segmentation are handled at initialization, while similarity ranking and style transfer operate asynchronously for responsive user feedback. The system is optimized for immersive performance, supporting high-resolution stereo output (5760×2880) at 30 FPS, compatible with commercial VR headsets.

5 Style encoding, similarity computation, and recommendations

The proposed framework extends *PanoStyle* [Tukur et al. 2023b] by introducing a structured methodology for extracting, comparing, and analyzing style codes from indoor panoramic scenes, enabling recommendation and clustering based on visual style.

Each panoramic indoor image is first processed using the *Multipanowise* transformer [Shah et al. 2024b] to obtain semantic segmentation maps and reflectance (albedo) information. These signals are used as input to a pre-trained *Panostyle* model that includes an *Adaptive Contextual Encoding (ACE)* module. The ACE module, built using SPADE-based conditional normalization, extracts latent style codes ϕ_c per semantic class c, robust to lighting variability.

The ACE module produces spatially localized embeddings for each object class using reflectance-aligned feature maps. Each latent code ϕ_c captures material and texture characteristics associated with the corresponding semantic region. These descriptors form the foundation for computing stylistic similarity across images.

The overall style of an image I is represented by concatenating the individual style codes of all semantic classes:

$$\Phi^I = [\phi_{c_1}, \phi_{c_2}, ..., \phi_{c_N}], \tag{1}$$

where missing classes are assigned zero vectors to ensure consistent dimensionality:

$$\phi_c = 0, \quad \text{if } c \notin C^I. \tag{2}$$

Style similarity is computed using a cosine-based metric restricted to shared semantic regions between two images:

$$\sigma(\Phi^A, \Phi^B) = \sum_{c \in C^A \cap C^B} \left(1 + \frac{\Phi_c^A \cdot \Phi_c^B}{\|\Phi_c^A\| \|\Phi_c^B\|}\right) \cdot \frac{1}{2|C^A \cap C^B|}. \tag{3}$$

This formulation ensures semantic alignment and perceptual relevance during similarity estimation.

Cosine similarity enables the retrieval of panoramas most similar to a query in style, and forms the basis for imposing an organization on the panorama database. For instance, the computed style codes and style similarity can support downstream tasks, including dimensionality reduction and style clustering. In particular, we can analyze the precomputed database by applying PCA and t-SNE to visualize stylistic relationships in a compact space and detect outliers. Moreover, unsupervised clustering algorithms such as K-Means, DBSCAN, and GMM can group panoramas based on shared stylistic traits. Fig. 2 summarizes the main steps of the similarity framework and the possible applications.

In particular, cosine similarity and unsupervised clustering can be exploited to provide style recommendations to the WebXR viewer each time a scene or an applied style changes. In particular, for broad navigation, the clustering imposed on the input images allows us to propose a variety of different styles by selecting representative template images from randomly selected clusters, excluding the cluster in which the current image and style fall. The selected clusters can also be organized by distance from the current one, using cosine similarity, thus permitting the organization of the thumbnails from the most similar to the most distant. This broad navigation approach can be combined with a local selection method for fine-tuning the selection by applying nearby styles. This can be achieved in two ways, depending on the desired scale of variation. The first method is to simply rank the clusters by similarity with the current ones using the cosine similarity from their centroids to the current styled panoramas, and select the first K nearest neighbors. The approach provides some variety, since we use clusters not including the current one, but also a good degree of similarity, since we remain in the neighborhood. For even finer search, it is also

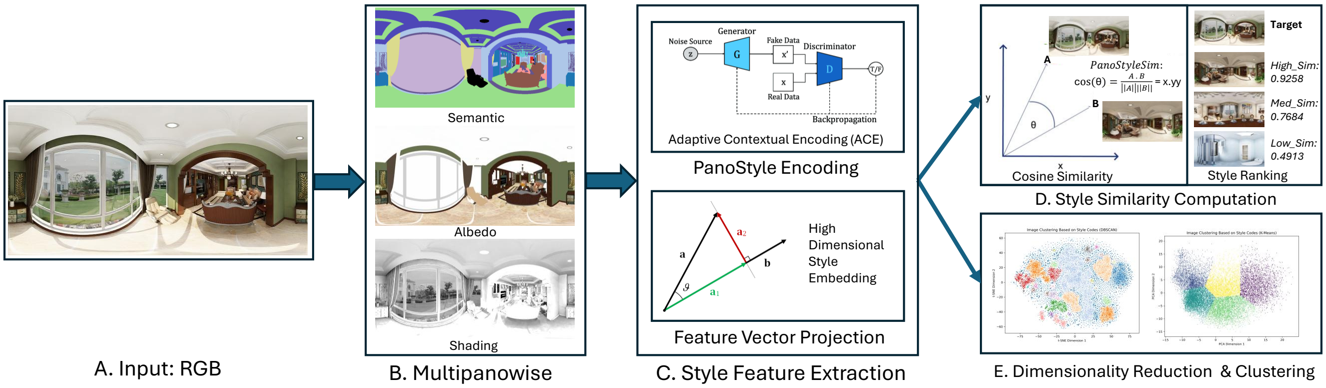

Figure 2: Framework for Panoramic Indoor Style Similarity. The framework consists of five main stages: (A) Input RGB, where a panoramic indoor image is provided as input; (B) Multipanowise Processing, which extracts semantic segmentation, albedo, and shading information; (C) Style Feature Extraction, utilizing Adaptive Contextual Encoding (ACE) to generate high-dimensional style embeddings; (D) Style Similarity Computation, where cosine similarity is used to rank images based on stylistic resemblance; and (E) Dimensionality Reduction & Clustering, employing PCA, t-SNE, and clustering algorithms (K-Means, DBSCAN, GMM) to structure and categorize style relationships. The framework facilitates robust style-based retrieval, ranking, and clustering of panoramic indoor environments.

possible to generate new style proposals by selecting styles close to the same image in different directions within the latent space. In our current prototype *WebXR* viewer and backend system, we have implemented only broad navigation (i.e., random selection from other clusters), leaving the integration of coarse and fine navigation to future work. The machinery to implement all these possibilities is already realized in the back-office components, and the structure of the styling space in our results section (Sec. 6).

6 Results

Our work combines a framework for extracting, organizing, and applying styles coming from panoramic images of indoor environments with a *WebXR* interface that supports immersive stereoscopic exploration, style selection, and application. In the following, we report on the results obtained with our proof of concept prototypes, first focusing on the analysis features (Sec. 6.1) and then on the immersive editing and exploration features (Sec. 6.2).

6.1 Style similarity

Performance metrics. The metric used for *Style Similarity Ranking* is the *Cosine Similarity Score*, which measures the angular distance between style feature vectors, ensuring that images with similar stylistic attributes yield higher similarity scores [Nguyen and Bai 2010]. This metric is fundamental to our ranking system, where a higher cosine similarity indicates greater stylistic resemblance.

Dimensionality reduction. To effectively structure the style feature space, we employ the following dimensionality reduction techniques.

- Principal Component Analysis (PCA): Reduces the high-dimensional style code representations to a lower-dimensional space, allowing for efficient visualization and interpretation of stylistic variations [Kurita 2021].
- t-Distributed Stochastic Neighbor Embedding (t-SNE): Provides a non-linear dimensionality reduction technique for visualizing high-dimensional style codes while preserving local similarities [Cieslak et al. 2020].

Clustering. We assess the ability of our framework to group stylistically similar images using the following techniques.

- K-Means Clustering: Assigns images to K distinct style clusters, ensuring that scenes with similar styles are grouped based on minimized intra-cluster variance [Sinaga and Yang 2020].
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN): Identifies clusters of arbitrary shape by considering the density distribution of style codes, effectively handling noise and outliers in the style space [Deng 2020].
- Gaussian Mixture Model (GMM): Models the distribution of style codes as a mixture of multiple Gaussian components. GMM allows for soft clustering where each image is associated with a probability of belonging to each cluster, capturing more nuanced stylistic relationships across indoor panoramas [Zhang et al. 2021].

These metrics ensure that our similarity ranking and clustering methods accurately capture stylistic relationships while maintaining computational efficiency.

Datasets. Our experimental evaluation utilizes the Structured3D dataset [Zheng et al. 2020], which comprises over 18,000 RGB-D images derived from 3,500 distinct indoor scenes representing diverse architectural designs. This dataset serves as an optimal benchmark for assessing the effectiveness of our proposed method, as it provides comprehensive ground truth annotations for panoramic indoor environments, including semantic segmentation, depth, surface normals, albedo, and RGB imagery. The dataset encompasses 41 region categories, mapped using NYU_v2 color palettes, covering key indoor elements such as walls, floors, cabinets, chairs,

Figure 3: Style similarity ranking results. Examples of style similarity comparison. The target image (left) is compared against three retrieved images with varying similarity levels: high, medium, and low. The *PanoStyleVR* metric is consistent with *SSIM* and *ArtFID*, effectively capturing stylistic resemblance. Higher *PanoStyleVR* values indicate stronger stylistic similarity, while lower values reflect greater divergence from the target image.

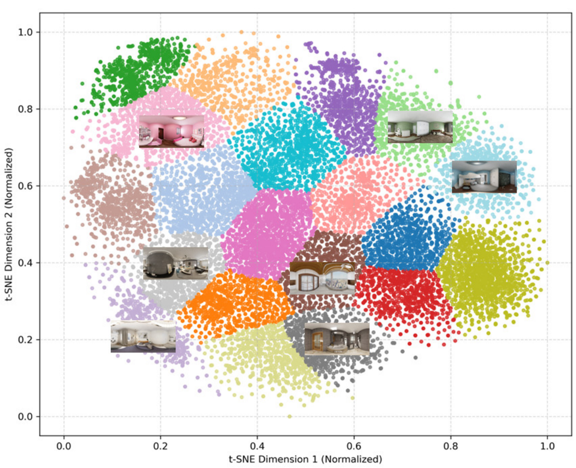

Figure 4: Clustering of Indoor Panoramic Scenes Based on Style Features Using Gaussian Mixture Model (GMM). The Figure illustrates the t-SNE projection of high-dimensional style feature vectors into a 2D space, where colors represent clusters identified by GMM (20 components). Representative panoramic images are overlaid to highlight the visual coherence within clusters. This visualization showcases the effectiveness of the PanoStyleVR pipeline in organizing large-scale indoor scenes based on latent stylistic attributes.

sofas, and tables. To facilitate efficient training and inference, we implemented a suite of pre-processing scripts: one to pre-compute shading images by integrating albedo and RGB frames, and another to extract labeled images from color-mapped semantic annotations. These preprocessing steps enhance the data preparation pipeline, ensuring robust input for our similarity ranking and clustering framework.

Qualitative and quantitative assessment. We conducted both qualitative and quantitative evaluations to assess the performance and perceptual consistency of the proposed *PanoStyleVR* framework. Initially, we visualize style similarity ranking results by comparing a target panoramic image with top-ranked retrieved scenes categorized into high, medium, and low similarity levels (cf. Fig. 3). These visual exemplars illustrate a strong alignment between *PanoStyleVR* scores and human perceptual judgments. Additionally, we incorporate the Structural Similarity Index (*SSIM*) and *ArtFID* as reference metrics to benchmark stylistic coherence, thereby confirming that *PanoStyleVR* scores remain consistent with established perceptual quality indicators.

We further evaluate the scalability of the similarity computation across a large dataset by analyzing the distribution of similarity scores. Moreover, clustering analysis was conducted on the Structured3D dataset [Zheng et al. 2020] using dimensionality reduction techniques, including PCA and t-SNE, complemented by unsupervised clustering via K-Means, DBSCAN, and Gaussian Mixture Models (GMM). Among these, t-SNE complemented by GMM produced the most coherent clustering outcomes, as illustrated in Fig. 4. To provide additional insight, we also present exemplar scenes from selected clusters in Fig. 5, demonstrating distinct interior design styles and structural patterns captured by the clustering process. Additionally, Fig. 6 presents a comparative visualization of PCA and t-SNE projections, highlighting how different dimensionality reduction techniques capture the underlying stylistic structure of the style code space. Furthermore, Fig. 7 compares the clustering performance of DBSCAN and K-Means on top of t-SNE: While DBSCAN showcases a significant number of outliers, K-Means is not able to capture clusters with irregular shapes. For these reasons, GMM was preferred for our image assessments (c.f. Fig. 4). These visualizations collectively validate the effectiveness of the *PanoStyleVR* framework in organizing indoor scenes based on latent style codes. Our results demonstrate that the proposed method robustly captures stylistic variations in panoramic indoor environments and facilitates data-driven scene classification and retrieval.

Limitations. While the proposed framework significantly advances style similarity ranking and clustering for panoramic indoor scenes, several limitations remain, necessitating further research:

Figure 5: Exemplar Scenes of Some Selected Clusters. The Figure presents exemplar scenes from five selected clusters from 20 clusters identified by GMM (Fig. 4), demonstrating distinct interior design styles and structural patterns captured by the clustering process. This further showcases the effectiveness of the PanoStyleVR pipeline in organizing large-scale indoor scenes based on latent stylistic attributes.

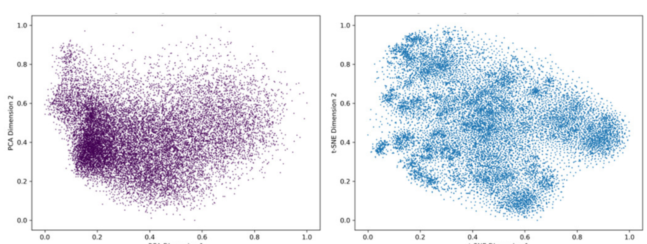

Figure 6: Visualization of Style Feature Space via Dimensionality Reduction. The figure presents the 2D projections of high-dimensional style codes using (left) Principal Component Analysis (PCA) and (right) t-Distributed Stochastic Neighbor Embedding (t-SNE). While PCA preserves global variance in the feature space, t-SNE more effectively captures local stylistic groupings, revealing the inherent structure and clusters of panoramic indoor scenes.

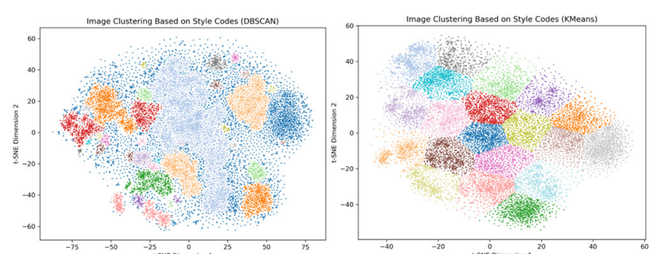

Figure 7: Comparative Visualization of Clustering Techniques Based on Style Codes. The figure illustrates example clustering results obtained using (left) DBSCAN and (right) K-Means, applied to the t-SNE projection of high-dimensional style codes.

- Style Code Generalization: The extracted NYU_v2-based style codes are optimized for indoor environments, limiting their applicability to outdoor or hybrid spaces. Non-standard layouts and custom interior designs may introduce inconsistencies, requiring adaptive feature extraction techniques for broader generalization.
- Computational Overhead and Scalability: High-dimensional style feature extraction incurs computational costs, particularly in large-scale datasets. While dimensionality reduction

(PCA, t-SNE) enhances efficiency, clustering techniques like DBSCAN, K-Means, and GMM may struggle with scalability, necessitating GPU acceleration or cloud-based processing for real-time applications.
- Sensitivity to Environmental Variations: The framework's semantic and geometric encoding can be affected by lighting conditions, occlusions, and reflections, leading to style inconsistencies. Highly reflective surfaces (e.g., glass, polished metals) and dynamic lighting introduce artifacts, requiring illumination-invariant feature representations.
- Limited Contextual Awareness: While the 41-class NYU_v2 palette effectively captures object-level style attributes, the framework lacks higher-order contextual understanding, such as user preferences, cultural influences, and design trends. Integrating multimodal learning (e.g., text-based adaptation, user-driven ranking) could enhance its contextual adaptability.
- Human Perception and Aesthetic Judgment: The current ranking and clustering mechanisms rely on numerical similarity metrics (cosine similarity, PCA, t-SNE, DBSCAN, K-Means, GMM), which may not fully align with human aesthetic perception. Future work should incorporate perceptual metrics, user studies, and qualitative evaluation to bridge computational style analysis with human-driven design intuition.

6.2 Immersive style transfer editing system

Figure 8: Immersive exploration and application of style transfer. We show frames from interactive sessions with our prototype immersive system. Casual users can explore stereoscopic indoor panoramic scenes and select styles from floating thumbnails to be applied to the scene they are currently exploring. See https://bit.ly/4donj0B for a full video.

We tested the immersive viewer and styling editor on various head-mounted display devices, including a Pico4, a Meta Quest 3, and a Google Cardboard. Here We report on experiments made on a Meta Quest 3, a headset with two 4.48-inch Fast-LCD displays, a global resolution of 4128 × 2208 pixels (equivalent to 2064 × 2208 pixels per screen), a pixel density (PPI) of 1218, a variable refresh rate of 72–120 Hz, and a diagonal Field of View (FOV) of 110 degrees.

To present an original or stylized image, only two MCOP panoramic images must be uploaded to the HMD, since the embedded client performs all the rendering computation and work. The client application is built on the *Three.js* framework using custom *WebGL* graphics components, and handles interaction through the *WebXR* APIs. Stereoscopic rendering is achieved by reading the sensor information to get head orientation, and separately rendering the left and right eye, each with its own MCOP image serving as a spherical environment. The viewer also allows switching from stereo to mono and, most importantly, makes it possible to constrain the up vector to stay aligned with the gravity direction during navigation. This latter option is important, as the original MCOP image is computed by assuming a gravity-aligned orientation. An informal test with 5 subjects indicated the preference for stereo mode and a locked vertical direction, as it provided a more immersive experience [Jashari et al. 2024]. For what concerns the interactivity, the immersive renderer can sustain fully interactive refresh rates (30fps for stereo), while the application of style transfer and the generation of MCOP stereo couples in the current implementation cannot be obtained in real time. For the preliminary system, we selected a set of styles according to the clusters showcased in Fig. 5 to be precomputed offline. Figures 1 and 8 show frames from interactive sessions with our prototype immersive system. Casual users can explore stereoscopic indoor panoramic scenes and select styles from floating thumbnails to be applied to the scene they are currently exploring. A full video showcasing examples of immersive application of panoramic style transfers is available at https://bit.ly/4donj0B.

7 Conclusion and future work

We have presented *PanoStyleVR*, an immersive web-based system for interactive style analysis and photorealistic adaptation in panoramic indoor scenes. The framework integrates a fully immersive *WebXR* interface for real-time, stereoscopic exploration, style selection, and application with a panoramic indoor style similarity framework that organizes sample scenes and provides style recommendations. The Panoramic Indoor Style Similarity framework represents a significant advancement in style-based ranking and clustering for panoramic indoor environments, offering impactful applications in virtual staging, real estate visualization, digital twins, and AR-driven interactive design. By leveraging NYU_v2-based style codes, dimensionality reduction, and unsupervised clustering techniques, the proposed methodology enables automated scene retrieval, classification, and content-aware editing.

Experiments on the Structured3D dataset validate the effectiveness of our perceptual similarity metric and the system prototype's ability to support immersive, style-driven scene editing.

We plan to extend our research and improve the current prototype, targeting both the *WebXR* viewer capabilities and the similarity framework. First of all, the current prototype is a proof-of-concept tested on a few scenes, where we could exploit several levels of precomputation. We plan to extend it to fully support dynamic exploration from fully novel scenes, with on-the-fly analysis, relighting, and stereo generation. We expect this to be doable, since the back-end neural models are very lean [Pintore et al. 2024b; Tukur et al. 2023b], and several parts of the analysis can be cached during interaction. Moreover, the current interactive system only supports recommendations of different styles. We plan to integrate the narrow search methods described in Sec. 5 by introducing a hierarchical style selection, where one can decide whether to expand the search, staying at the same level, or narrow it by stepping down. In this context, we also plan to perform a human-centric perceptual evaluation to bridge the gap between automated similarity ranking and human aesthetic judgment and to evaluate the effectiveness of style navigation interfaces. Expanding the framework to diverse architectural and cultural datasets beyond NYU_v2 could also improve generalization across varying indoor styles and layouts.

By addressing these research directions, we aim to evolve the framework into a more robust, scalable, and perceptually-aware AI-driven tool, advancing applications in interactive virtual environments and automated interior design. In particular, the proposed immersive style transfer framework has broad applicability across multiple domains, particularly in virtual staging [Shah et al. 2024a; Tukur et al. 2024], real estate visualization, digital twins, and immersive scene exploration.

Acknowledgments

This publication was made possible by NPRP-Standard (NPRP-S) 14th Cycle grant 0403-210132 AIN2 from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work and are solely the responsibility of the authors.

References

Cristhian A. Aguilera, Francisco J. Aguilera, Angel D. Sappa, Cristhian Aguilera, and Ricardo Toledo. 2016. Learning Cross-Spectral Similarity Measures With Deep Convolutional Neural Networks. In *Proc. CVPR Workshops*. IEEE, Las Vegas, Nevada, 267–275.

Lucas F Buzuti and Carlos E Thomaz. 2023. Fréchet AutoEncoder distance: a new approach for evaluation of generative adversarial networks. *Computer Vision and Image Understanding* 235 (2023), 103768.

Matthew C Cieslak, Ann M Castelfranco, Vittoria Roncalli, Petra H Lenz, and Daniel K Hartline. 2020. t-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Marine genomics* 51 (2020), 100723.

Lavinia De Divitiis, Federico Becattini, Claudio Baecchi, and Alberto Del Bimbo. 2023. Disentangling Features for Fashion Recommendation. *ACM Trans. Multimedia Comput. Commun. Appl.* 19, 1s, Article 39 (Jan. 2023), 21 pages.

Yashar Deldjoo, Fatemeh Nazary, Arnau Ramisa, Julian McAuley, Giovanni Pellegrini, Alejandro Bellogin, and Tommaso Di Noia. 2023. A Review of Modern Fashion Recommender Systems. *ACM Comput. Surv.* 56, 4, Article 87 (Oct. 2023), 37 pages.

Dingsheng Deng. 2020. DBSCAN clustering algorithm based on density. In *Proc. IFEA*. IEEE, IEEE, Hennai, India, 949–953.

Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. 2020. Image quality assessment: Unifying structure and texture similarity. *IEEE TPAMI* 44, 5 (2020), 2567–2581.

Elena Garces, Aseem Agarwala, Diego Gutierrez, and Aaron Hertzmann. 2014. A similarity measure for illustration style. *ACM TOG* 33, 4, Article 93 (July 2014), 9 pages.

Enrico Gobbetti, Giovanni Pintore, and Marco Agus. 2024. Automatic 3D modeling and exploration of indoor structures from panoramic imagery. In *Proc. SIGGRAPH Asia Courses*. ACM, Tokyo,Japan, Article 1, 9 pages.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *NeurIPS* 27 (2014), 1–9.

Kshitij Gulati, Gaurav Verma, Mukesh Mohania, and Ashish Kundu. 2023. BeautifAI-Personalised Occasion-based Makeup Recommendation. In *Asian Conference on Machine Learning*. PMLR, PMLR, Istanbul,Turkey, 407–419.

Peter Hedman and Johannes Kopf. 2018. Instant 3D photography. *ACM TOG* 37, 4 (2018), 1–12.

Jingwei Huang, Zhili Chen, Duygu Ceylan, and Hailin Jin. 2017. 6-DOF VR videos with a single 360-camera. In *Proc. IEEE VR*. IEEE, Los Angeles, California, 37–44.

Murium Iqbal, Kamelia Aryafar, and Timothy Anderton. 2019. Style conditioned recommendations. In *Proc. RecSys* (Copenhagen, Denmark). ACM, New York, NY, USA, 128–136.

Sara Jashari, Muhammad Tukur, Yehia Boraey, Mahmood Alzubaidi, Giovanni Pintore, Enrico Gobbetti, Alberto Jaspe Villanueva, Jens Schneider, Noora Fetais, and Marco Agus. 2024. Evaluating AI-based static stereoscopic rendering of indoor panoramic scenes. In *STAG: Smart Tools and Applications in Graphics*. Eurographics Association, Verona,Italy, 9 pages.

Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. 2019. Neural style transfer: A review. *IEEE TVCG* 26, 11 (2019), 3365–3385.

Ilham Kachbal, Said El Abdellaoui, and Khadija Arhid. 2024. Revolutionizing Fashion Recommendations: A Deep Dive into Deep Learning-based Recommender Systems. In *Proc. NISS* (Meknes, AA, Morocco). ACM, New York, NY, USA, Article 1, 8 pages.

Jinsung Kim and Jin-Kook Lee. 2020. Stochastic detection of interior design styles using a deep-learning model for reference images. *Applied Sciences* 10, 20 (2020), 7299.

Hyeyoung Ko, Suyeon Lee, Yoonseo Park, and Anna Choi. 2022. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics* 11, 1 (2022), 141:1–141::48.

Takio Kurita. 2021. Principal component analysis (PCA). In *Computer vision: a reference guide*. Springer, Berlin,Germany, 1013–1016.

Qinbo Li and Nima Khademi Kalantari. 2020. Synthesizing Light Field From a Single Image with Variable MPI and Two Network Fusion. *ACM TOG* 39, 6 (2020), 229:1–229:10.

Tianqiang Liu, Aaron Hertzmann, Wilmot Li, and Thomas Funkhouser. 2015. Style compatibility for 3D furniture models. *ACM TOG* 34, 4, Article 85 (July 2015), 9 pages.

Zhaoliang Lun, Evangelos Kalogerakis, and Alla Sheffer. 2015. Elements of style: learning perceptual shape style similarity. *ACM TOG* 34, 4, Article 84 (July 2015), 14 pages.

Bicheng Luo, Feng Xu, Christian Richardt, and Jun-Hai Yong. 2018. Parallax360: Stereoscopic 360 scene representation for head-motion parallax. *IEEE TVCG* 24, 4 (2018), 1545–1553. Proc. IEEE VR.

Hui Mao, Ming Cheung, and James She. 2017. DeepArt: Learning Joint Representations of Visual Arts. In *Proc. MM* (Mountain View, California, USA). ACM, New York, NY, USA, 1183–1191.

Kevin Matzen, Michael F. Cohen, Bryce Evans, Johannes Kopf, and Richard Szeliski. 2017. Low-cost 360 Stereo Photography and Video Capture. *ACM TOG* 36, 4 (2017), 148:1–148:12.

Muzammal Naseer, Salman Khan, and Fatih Porikli. 2019. Indoor Scene Understanding in 2.5/3D for Autonomous Agents: A Survey. *IEEE Access* 7 (2019), 1859–1887.

Hieu V Nguyen and Li Bai. 2010. Cosine similarity metric learning for face verification. In *Asian conference on computer vision*. Springer, Springer, Queenstown,New Zealand, 709–720.

Artem Obukhov and Mikhail Krasnyanskiy. 2020. Quality assessment method for GAN based on modified metrics inception score and Fréchet inception distance. In *Proc. CMSS Vol. 1 4*. Springer, Springer, Online, Virtual, 102–114.

Giovanni Pintore, Marco Agus, Eva Almansa, and Enrico Gobbetti. 2022. Instant Automatic Emptying of Panoramic Indoor Scenes. *IEEE Transactions on Visualization and Computer Graphics* 0, 0 (Nov. 2022), 1–11. doi:10.1109/tvcg.2022.3202999

Giovanni Pintore, Marco Agus, Alberto Signoroni, and Enrico Gobbetti. 2024a. DDD: Deep indoor panoramic Depth estimation with Density maps consistency. In *STAG: Smart Tools and Applications in Graphics*. Eurographics, Verona,Italy, 10. doi:10.2312/stag.20241336

Giovanni Pintore, Fabio Bettio, Marco Agus, and Enrico Gobbetti. 2023. Deep scene synthesis of Atlanta-world interiors from a single omnidirectional image. *IEEE TVCG* 29 (November 2023), 1–10.

Giovanni Pintore, Fabio Ganovelli, Alberto Jaspe Villanueva, and Enrico Gobbetti. 2019. Automatic modeling of cluttered multi-room floor plans from panoramic images. *Computers Graphics Forum* 38, 7 (2019), 347–358.

Giovanni Pintore, Alberto Jaspe-Villanueva, Markus Hadwiger, Jens Schneider, Marco Agus, Fabio Marton, Fabio Bettio, and Enrico Gobbetti. 2024b. Deep synthesis and exploration of omnidirectional stereoscopic environments from a single surround-view panoramic image. *Computers & Graphics* 119 (2024), 103907.

Giovanni Pintore, Claudio Mura, Fabio Ganovelli, Lizeth Fuentes-Perez, Renato Pajarola, and Enrico Gobbetti. 2020. State-of-the-art in Automatic 3D Reconstruction of Structured Indoor Environments. *Comput. Graph. Forum* 39, 2 (2020), 667–699.

Hamed Qazanfari, Mohammad M AlyanNezhadi, and Zohreh Nozari Khoshdaregi. 2023. Advancements in content-based image retrieval: A comprehensive survey of relevance feedback techniques. *arXiv preprint arXiv:2312.10089* 0, 0 (2023), 10 pages.

Dan Ruta, Saeid Motiian, Baldo Faieta, Zhe Lin, Hailin Jin, Alex Filipkowski, Andrew Gilbert, and John Collomosse. 2021. ALADIN: All Layer Adaptive Instance Normalization for Fine-Grained Style Similarity. In *Proc. ICCV*. IEEE, Montreal, Canada, 11926–11935.

Uzair Shah, Sara Jashari, Muhammad Tukur, Giovanni Pintore, Enrico Gobbetti, Jens Schneider, and Marco Agus. 2024a. VISPI: Virtual Staging Pipeline for Single Indoor Panoramic Images. In *STAG: Smart Tools and Applications in Graphics*. Eurographics, Verona,Italy, 10 pages. doi:10.2312/stag.20241334

Uzair Shah, Muhammad Tukur, Mahmood Alzubaidi, Giovanni Pintore, Enrico Gobbetti, Mowafa Househ, Jens Schneider, and Marco Agus. 2024b. MultiPanoWise: Holistic Deep Architecture for Multi-task Dense Prediction from a Single Panoramic Image. In *Proc. CVPR Workshops*. IEEE, Seattle, USA, 1311–1321.

Raksha Sharma. 2024. *Virtual Staging for Real Estate Market Research Report 2033*. DataIntelo, USA.

H.R. Sheikh and A.C. Bovik. 2006. Image information and visual quality. *IEEE TIP* 15, 2 (2006), 430–444.

Xi Shen, Alexei A. Efros, and Mathieu Aubry. 2019. Discovering Visual Patterns in Art Collections With Spatially-Consistent Feature Learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Long Beach, California, 9270–9279. doi:10.1109/CVPR.2019.00950

Jia Wei Sii and Chee Seng Chan. 2025. Gorgeous: Creating narrative-driven makeup ideas via image prompts. *Multimedia Tools and Applications* 0, 0 (2025), 1–22.

Kristina P Sinaga and Miin-Shen Yang. 2020. Unsupervised K-means clustering algorithm. *IEEE access* 8 (2020), 80716–80727.

Gowthami Somepalli, Anubhav Gupta, Kamal Gupta, Shramay Palta, Micah Goldblum, Jonas Geiping, Abhinav Shrivastava, and Tom Goldstein. 2024. Measuring Style Similarity in Diffusion Models. *CoRR* abs/2404.01292, 0 (2024), 10 pages.

Gowthami Somepalli, Anubhav Gupta, Kamal Gupta, Shramay Palta, Micah Goldblum, Jonas Geiping, Abhinav Shrivastava, and Tom Goldstein. 2025. Investigating Style Similarity in Diffusion Models. In *Proc. ECCV*, Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (Eds.). Springer Nature Switzerland, Cham, 143–160.

Richard Tucker and Noah Snavely. 2020. Single-View View Synthesis With Multiplane Images. In *Proc. CVPR*. IEEE, Vancouver, Canada, 548–557.

Muhammad Tukur, Yehia Boraey, Sara Jashari, Alberto Jaspe Villanueva, Uzair Shah, Mahmood Al-Zubaidi, Giovanni Pintore, Enrico Gobbetti, Jens Schneider, Marco Agus, et al. 2024. Virtual Staging Technologies for the Metaverse. In *Proc. iMeta*. IEEE, IEEE, Dubai,UAE, 206–213.

Muhammad Tukur, Giovanni Pintore, Enrico Gobbetti, Jens Schneider, and Marco Agus. 2023a. SPIDER: A framework for processing, editing and presenting immersive high-resolution spherical indoor scenes. *Graphical Models* 128 (2023), 101182.

Muhammad Tukur, A Ur Rehman, Giovanni Pintore, Enrico Gobbetti, Jens Schneider, and Marco Agus. 2023b. PanoStyle: Semantic, geometry-aware and shading independent photorealistic style transfer for indoor panoramic scenes. In *Proc. ICCV*. IEEE, Paris, France, 1553–1564.

John Waidhofer, Richa Gadgil, Anthony Dickson, Stefanie Zollmann, and Jonathan Ventura. 2022. PanoSynthVR: Toward Light-weight 360-Degree View Synthesis from a Single Panoramic Input. In *Proc. ISMAR*. IEEE, Singapore, 584–592.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP* 13, 4 (2004), 600–612.

Matthias Wright and Björn Ommer. 2022. Artfid: Quantitative evaluation of neural style transfer. In *DAGM German Conference on Pattern Recognition*. Springer, Springer, Konstanz, Germany, 560–576.

Chuyi Xiong, Ka Shing Cheung, Deborah Susan Levy, and Michael Allen. 2024. The effect of virtual reality on the marketing of residential property. *Housing Studies* 39, 3 (2024), 671–694.

Jiale Xu, Jia Zheng, Yanyu Xu, Rui Tang, and Shenghua Gao. 2021. Layout-guided novel view synthesis from a single indoor panorama. In *Proc. CVPR*. IEEE, Nashville, Tennessee, 16438–16447.

Yao Xu, Min Xia, Kai Hu, Siyi Zhou, and Liguo Weng. 2025. Style Transfer Review: Traditional Machine Learning to Deep Learning. *Information* 16, 2 (2025), 157:1–157:24.

Yang Yang, Shi Jin, Ruiyang Liu, , and Jingyi Yu. 2018. Automatic 3D Indoor Scene Modeling From Single Panorama. In *Proc. CVPR*. IEEE, Salt Lake City, Utah, 3926–3934.

Wujian Ye, Yue Wang, Yijun Liu, Wenjie Lin, and Xin Xiang. 2025. Panoramic Arbitrary Style Transfer with Deformable Distortion Constraints. *J Vis. Comm. and Im. Rep.* 106 (2025), 104344.

Yi Zhang, Miaomiao Li, Siwei Wang, Sisi Dai, Lei Luo, En Zhu, Huiying Xu, Xinzhong Zhu, Chaoyun Yao, and Haoran Zhou. 2021. Gaussian mixture model clustering with incomplete data. *ACM TOMM* 17, 1s (2021), 1–14.

Yiwei Zhang and Toshihiko Yamasaki. 2021. Style-Aware Image Recommendation for Social Media Marketing. In *Proc. ACM MM* (Virtual Event, China). ACM, New York, NY, USA, 3106–3114.

Yiqun Zhao, Zibo Zhao, Jing Li, Sixun Dong, and Shenghua Gao. 2024. RoomDesigner: Encoding Anchor-latents for Style-consistent and Shape-compatible Indoor Scene Generation. In *Proc. 3DV*. IEEE, Davos, Switzerland, 1413–1423.

Jia Zheng, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. 2020. Structured3D: A Large Photo-realistic Dataset for Structured 3D Modeling. In *Proc. ECCV*. Springer, Glasgow, UK, 519–535.

Tiancheng Zhi, Bowei Chen, Ivaylo Boyadzhiev, Sing Bing Kang, Martial Hebert, and Srinivasa G Narasimhan. 2022. Semantically supervised appearance decomposition for virtual staging from a single panorama. *ACM TOG* 41, 4 (2022), 1–15.

Xiaotong Zhou, Yuhui Zheng, and Junming Yang. 2025. Bridging the metrics gap in image style transfer: A comprehensive survey of models and criteria. *Neurocomputing* 624 (2025), 129430.

Yufeng Zhou, Mei Yu, Hualin Ma, Hua Shao, and Gangyi Jiang. 2018. Weighted-to-Spherically-Uniform SSIM Objective Quality Evaluation for Panoramic Video. In *Proc. ICSP*. IEEE, Beijing, China, 54–57.

Jana Zujovic, Thrasyvoulos N Pappas, and David L Neuhoff. 2013. Structural texture similarity metrics for image analysis and retrieval. *IEEE TIP* 22, 7 (2013), 2545–2558.