Interior Design AI for Style Transformation with Structural Preservation

Alfonso Mateos Vicente, and Emiliano Pizaña Vela
{alfonso.mateos-vicente, emiliano.pizana-vela}@polytechnique.edu

***Abstract*—This report presents an advanced interior design AI system that combines generative AI for style transformation with structural preservation capabilities. By integrating depth map estimation, semantic segmentation, and diffusion-based image generation, our system adapts to existing interior spaces and generates realistic design variations while maintaining architectural integrity. Using DPT-large for depth estimation and Stable Diffusion XL with ControlNet for image generation, the pipeline creates coherent spatial transformations across multiple interior design styles. Our primary focus is the development of a style-guided transformation pipeline that respects spatial constraints while enabling creative redesign. Quantitative evaluation using metrics including Edge IoU, PSNR, SSIM, and LPIPS confirms our approach's effectiveness, with optimal performance at guidance scale 10 and style strength 0.4-0.5. Experiments on a dataset of 50 high-definition interior images demonstrate successful style transformations, particularly for bohemian designs where significant aesthetic changes emerge while respecting the original architectural environment. Additionally, we explore basic 3D visualization capabilities to complement the 2D transformations. This implementation represents a practical application of computer vision and generative AI techniques to interior design, making sophisticated style transformation accessible through an intuitive interface.**

I. Introduction

Interior design traditionally requires specialized skills and significant time investment, often limiting accessibility and iterative exploration. As design software evolves, there remains a gap between visualization tools and generative capabilities that can adapt to existing spaces while maintaining structural coherence.

This project introduces a comprehensive AI framework for interior design that combines depth map processing and controllable image generation. By leveraging diffusion models and neural networks, our system analyzes existing interiors, extracts structural information, and generates redesigned variations that maintain spatial consistency while transforming aesthetic elements.

Our framework operates through multiple integrated modules: image processing to extract depth and segmentation data, and controlled image generation to transform spaces according to user-specified style parameters. This approach enables users to visualize complete room transformations with structural fidelity. As a secondary capability, the system also provides basic 3D visualization to complement the 2D transformations.

The key contributions of this work include:

- An end-to-end pipeline integrating depth estimation, structural analysis, and generative design into a cohesive system
- A comprehensive style parameterization system implementing six design aesthetics with controlled transformations
- A robust evaluation framework with multiple perceptual and structural metrics for assessing transformation quality
- Optimized implementation for practical real-world use with graceful fallback mechanisms

Unlike previous approaches that focus on either visualization or generation separately, our system bridges both capabilities, allowing users to transform interior spaces with architectural awareness through a single integrated interface.

II. Methodology

A. System Architecture

Our interior design AI system consists of four primary modules working in concert to transform and visualize interior spaces:

- **Image Processor**: Extracts depth maps, segmentation masks, and room analysis from input images. Using DPT-large for depth estimation, SegFormer-B5 for semantic segmentation, BLIP for scene understanding, and CLIPSeg for object detection, this module creates a comprehensive analysis of the input space including structural elements, design style, and object placement.
- **Interior Reimaginer**: Transforms spaces using controlled diffusion models based on style parameters. By leveraging Stable Diffusion XL (with approximately 300M parameters) and ControlNet conditioning on both depth maps and edge detection, this module preserves structural elements while applying consistent style transformations.
- **Depth Reconstructor**: Provides basic 3D visualization capabilities as a complementary feature. This module implements simple point cloud generation from depth maps to offer spatial context for the redesigned interiors.
- **User Interface**: Gradio-based interactive interface that ties the system together, allowing users to upload images, select styles, adjust transformation parameters, and explore design variations. The interface provides immediate feedback through a side-by-side comparison of original and transformed spaces.

The integration of these modules enables a complete pipeline from input image to redesigned spaces, with particular emphasis on the style transformation capabilities while respecting structural constraints.

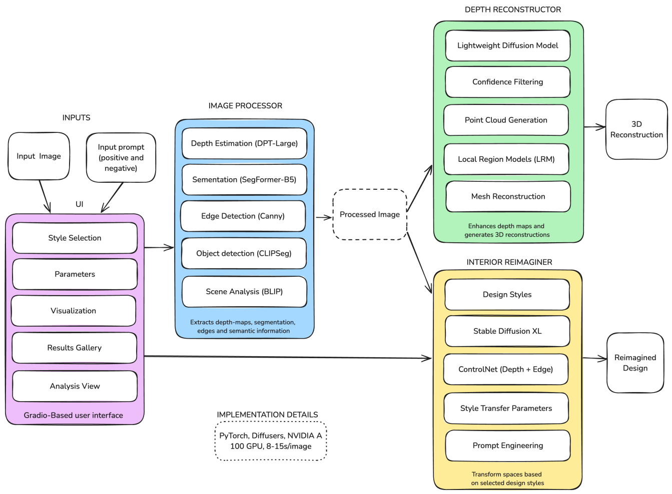

Fig. 1. Visualisation of the System Architecture.

B. Image Processing and Analysis

The ImageProcessor module serves as the foundation for spatial understanding, extracting multiple data representations from input images:

- **Depth Estimation**: Using DPT-large model to generate high-quality depth maps with fine-grained distance information
- **Semantic Segmentation**: Identifying structural elements such as walls, floors, ceilings, and furniture using SegFormer-B5, enabling targeted modifications
- **Edge Detection**: Capturing structural boundaries using Canny edge detection for preservation during transformation
- **Object Detection**: Generating masks for specific elements (furniture, decor) using CLIPSeg, allowing text-guided segmentation
- **Scene Understanding**: Automatic captioning with BLIP to identify design styles, materials, colors, and spatial arrangements

The processed data is encapsulated in a ProcessedImage container that provides structured input for the design generation pipeline. This multi-faceted analysis enables the system to understand both the aesthetic and structural elements of the space, which is crucial for maintaining architectural integrity during transformation.

C. Interior Reimagining with Diffusion Models

The InteriorReimaginer module—the core of our system—generates transformed designs using state-of-the-art diffusion models with several key innovations:

- **Stable Diffusion XL**: We employ the SDXL base model (approximately 300M parameters) for high-quality image generation, providing the foundation for realistic interior transformations.
- **ControlNet Conditioning**: Our system uses depth maps and edge detection as control signals, preserving structural elements during transformation. This dual conditioning ensures that architectural features remain consistent while aesthetic elements are transformed.
- **Style Parameterization**: We formalized six interior design styles as structured data classes with specific prompt modifiers, negative modifiers, and color palettes for consistent aesthetic transformations.
- **Transformation Controls**: The system provides fine-grained control over the transformation process through parameters like style strength (0.2-0.9) and guidance scale (1.0-15.0), allowing users to balance creativity and structural fidelity.

The reimagining process uses a combination of prompt engineering and control signals according to the formula:

$$I_{\text{new}} = G(I_{\text{original}}, \text{depth}, \text{edges}, \text{prompt}, \alpha) \tag{1}$$

where G is the generative model, and α controls transformation strength. This approach ensures that transformations respect the physical constraints of the space while allowing creative freedom in design elements.

What makes this approach particularly effective is how it handles the hierarchical nature of design elements. Architectural features receive the highest priority (preserved through ControlNet), followed by major furniture pieces (respected but potentially restyled), with decorative elements given the most creative freedom. This hierarchy ensures that transformations remain physically plausible while embracing style-appropriate aesthetics.

III. Implementation Details

A. Computational Infrastructure

Our implementation was developed and evaluated using the following infrastructure:

- **Computing Environment**: Google Colab with NVIDIA A100 GPU (40GB VRAM)
- **Software Stack**: Python 3.9+, PyTorch 2.0.1, Diffusers 0.19.3, Transformers 4.30.2, Gradio 3.32.0
- **Inference Parameters**: 30-50 diffusion steps, guidance scale 5-15, batch size 1-4
- **Processing Time**: 8-15 seconds per image on NVIDIA A100

The system includes optimizations for memory efficiency, including attention slicing and xformers support when available.

B. Neural Network Models

Our implementation integrates multiple specialized models:

- **Depth Estimation**: Intel/DPT-large for high-quality initial depth maps with fine-grained distance information
- **Segmentation**: NVIDIA SegFormer-B5 for semantic understanding with 150 ADE20K classes, providing detailed scene parsing
- **Image Captioning**: BLIP for automatic scene analysis, extracting style, colors, materials, and spatial arrangements

- **Object Segmentation**: CLIPSeg for text-guided object masks, enabling targeted segmentation of specific elements
- **Image Generation**: StableDiffusionXL with ControlNet for guided design, using 640×640 resolution with VAE 80×80×4 latent space

For efficient inference, we implement model weight caching and selective loading, reducing redundant network requests and optimizing memory usage. The system checks for cached weights before downloading, improving startup performance especially in environments with limited connectivity.

C. Design Style Parameterization

We formalized design styles as dataclasses with specific parameters:

- **Minimalist**: Clean lines, minimal decoration, functional furniture, neutral colors with prompt modifiers including ”clean lines,” ”uncluttered,” and negative modifiers ”ornate,” ”cluttered”
- **Scandinavian**: Light, airy spaces with wooden elements and cozy textiles with prompt modifiers ”light wood,” ”white walls,” and negative modifiers ”dark colors,” ”heavy furniture”
- **Industrial**: Raw materials, exposed structures, and vintage elements with prompt modifiers ”exposed brick,” ”metal fixtures,” and negative modifiers ”colorful,” ”delicate”
- **Mid-Century Modern**: Retro furniture, clean lines, and bold accent colors with prompt modifiers ”retro furniture,” ”organic curves,” and negative modifiers ”rustic,” ”traditional”
- **Bohemian**: Eclectic, colorful spaces with mixed patterns and textures with prompt modifiers ”colorful textiles,” ”mixed patterns,” and negative modifiers ”minimal,” ”monochrome”
- **Traditional**: Classic furniture, rich colors, and elegant details with prompt modifiers ”classic furniture,” ”elegant details,” and negative modifiers ”modern,” ”industrial”

Each style includes a comprehensive set of prompt modifiers, negative modifiers, and RGB color palettes that guide the generation process. This structured approach ensures consistent aesthetic transformations and enables meaningful comparison between styles.

The prompt construction process combines these elements, generating prompts like: ”Interior design: scandinavian, light wood, white walls, cozy textiles, natural light, functional design, photorealistic, high quality,” with corresponding negative prompts: ”low quality, blurry, distorted proportions, dark colors, heavy furniture, ornate decoration.”

During development, we discovered that certain styles present unique challenges. Industrial style, with its emphasis on exposed structural elements, sometimes conflicts with depth-conditioning that attempts to maintain smooth wall surfaces. Bohemian style, with its embrace of controlled chaos, occasionally challenges our structural preservation mechanisms. These edge cases required careful balancing of parameters and additional style-specific guidance to resolve effectively.

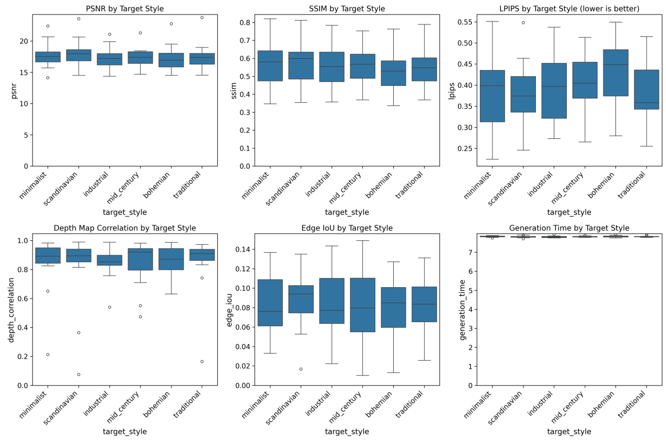

Fig. 2. Quantitative analysis by style.

IV. Experimental Results

A. Dataset and Evaluation Methodology

We conducted a thorough evaluation using:

- **Dataset**: 50 high-definition interior images from the Pexels database, covering a diverse range of rooms and styles
- **Transformation Variations**: Each image was transformed using all six design styles with varying parameters
- **Parameter Sweep**: Guidance scale (5-15), style strength (0.2-0.9), and diffusion steps (30-50)
- **Metrics**: PSNR, SSIM, LPIPS perceptual distance, Edge IoU, and depth correlation

For Edge IoU, we used the formula $\text{IoU} = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|}$ to quantify how well architectural boundaries are maintained, where E_1 and E_2 represent the edge maps of the original and transformed images. This metric was particularly important for evaluating structural preservation during style transfer.

B. Quantitative Results

Our evaluation revealed several key insights about the system’s performance:

- **Optimal Guidance Scale**: Best results achieved at guidance scale 10, balancing creative transformation with structural fidelity as measured across LPIPS, PSNR, SSIM, and Edge IoU metrics
- **Style Strength Impact**: Values between 0.4-0.5 offered the ideal balance between design changes and spatial consistency, with structural integrity degrading rapidly above 0.7
- **Style-Dependent Performance**: Scandinavian and traditional styles achieved better structural preservation (LPSIS 0.37 and 0.35) than bohemian style (LPSIS 0.45), logically reflecting the greater deviation from conventional design

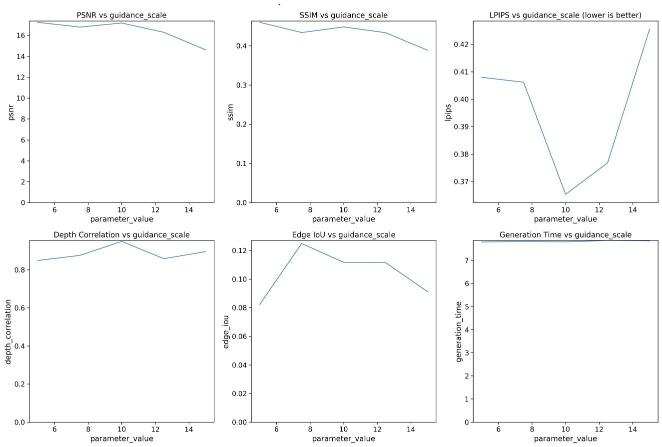

Fig. 3. Ablation study for the Guidance Scale parameter.

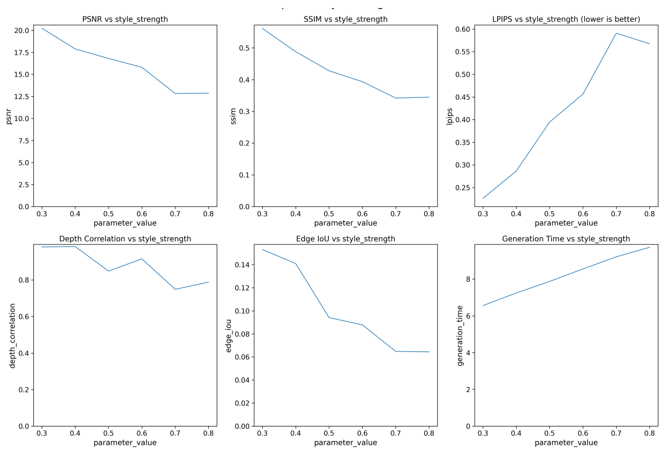

Fig. 4. Ablation study for the Style Strength parameter.

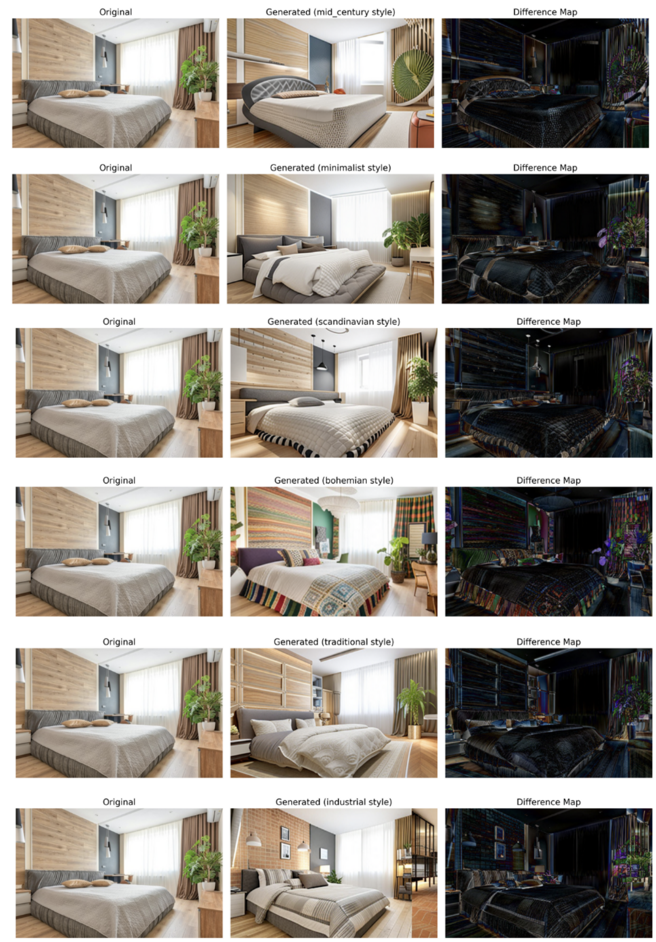

Fig. 5. Examples of generated images for qualitative analysis

Figure 3 demonstrates these metrics as a function of guidance scale, showing optimal performance around guidance scale 10. Figure 4 illustrates how style strength affects the balance between creativity and structural preservation, with values between 0.4-0.5 offering the best compromise.

C. Qualitative Analysis

Figure 5 showcases examples of different style transformations applied to the same space, demonstrating the system's ability to create diverse yet spatially consistent designs. The transformations maintain structural elements like walls, windows, and room layout while changing decorative elements, colors, and furniture styles according to the selected aesthetic.

What's particularly noteworthy is how the transformations handle specific design elements. Floor materials change appropriately—warm woods for Scandinavian, patterned rugs for bohemian. Wall treatments adapt to each style—clean white for minimalist, exposed brick for industrial, textured and colorful for bohemian. Furniture pieces maintain their approximate position and function but transform in appearance to match each style's aesthetic.

The study also revealed fascinating edge cases. Rooms with highly distinctive architectural features (exposed beams, unusual window shapes, built-in elements) maintained these characteristics across style transformations. Spaces with strong existing style elements sometimes showed ”style resistance”—requiring higher transformation strength to overcome the visual dominance of existing design choices.

V. Additional Capabilities: 3D Visualization

As a complementary feature to our main style transformation focus, we implemented basic 3D visualization capabilities:

- **Point Cloud Generation**: Converting depth maps to simple 3D point clouds for spatial context
- **Basic Visualization**: Rendering techniques to view the transformed space from different angles

While not the primary focus of our research, these capabilities provide additional context for understanding the spatial implications of design transformations. The 3D visualization complements the main style transformation pipeline by offering a different perspective on the redesigned spaces.

VI. Limitations and Future Work

A. Current Limitations

Our system faces several challenges:

- **Computational Requirements**: High-end GPU (NVIDIA A100 or equivalent) needed for optimal real-time performance, though the system functions with degraded performance on consumer hardware
- **Single-View Limitations**: Incomplete spatial understanding from one image, leading to potential inconsistencies in areas not visible in the input
- **Lighting Consistency**: Difficulty maintaining natural lighting models across style transfers, particularly for styles with significantly different lighting aesthetics
- **Material Specification**: Limited granular control over specific materials in generation, sometimes producing textures inconsistent with the selected style

B. Future Research Directions

Promising avenues for improvement include:

- **Multi-View Processing**: Incorporating multiple viewpoints for more complete spatial understanding
- **Material-Specific Control**: Enhancing control over surface materials and textures through additional conditioning signals
- **Lighting Simulation**: Physics-based rendering for accurate lighting visualization that responds to architectural features
- **Furniture Placement Optimization**: Automated arrangement suggestions based on room function and design principles
- **Enhanced 3D Capabilities**: More sophisticated 3D reconstruction and visualization for better spatial understanding

VII. Conclusion

This project began with a simple observation: people want to explore design possibilities for their existing spaces, but most lack either the visualization skills or the design expertise to do so effectively. By bringing together advances in depth estimation and generative AI, we've created a system that helps bridge this gap—making interior design exploration more accessible while respecting physical reality.

What distinguishes our approach is its integration of structural awareness with style transformation. Rather than treating these as separate problems, we've developed a unified framework that allows users to reimagine their spaces visually while respecting architectural constraints. This dual perspective is crucial for practical interior design, where aesthetic choices must work within architectural constraints.

Our experimental results demonstrate that this integrated approach works. The system successfully transforms interior spaces across multiple design styles while maintaining structural integrity, as confirmed by both quantitative metrics and qualitative assessment. Importantly, the entire pipeline functions within a user-friendly interface that makes these advanced capabilities accessible to non-experts.

The key advantages of our approach include:

- A comprehensive pipeline that takes users from a single photo to multiple design variations with minimal technical expertise required
- Depth-enhanced understanding that respects the physical reality of spaces during creative transformation
- Structured style parameterization that captures the multifaceted nature of interior design aesthetics
- Efficient implementation that balances quality with performance across different hardware environments

Perhaps most importantly, this system represents a step toward democratizing design exploration. By making it easier for people to visualize possibilities for their spaces, we potentially expand access to quality interior design—allowing more people to create environments that reflect their aesthetic preferences while respecting architectural reality.

References

[1] Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision transformers for dense prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12179-12188).

[2] Podell, D., et al. (2023). SDXL: Improving latent diffusion models for high-resolution image synthesis. In International Conference on Learning Representations.

[3] Zhang, L., Agrawala, M., & Durand, F. (2023). ControlNet: Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[4] Li, J., Li, D., Xiong, C., & Hoi, S. (2022). BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In International Conference on Machine Learning (pp. 12888-12900).

[5] Lüddecke, T., & Ecker, A. (2022). Image segmentation using text and image prompts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7086-7096).