

# INFO8010: Interior Design Style Transfer

Marie Goffin<sup>1</sup> and Thomas Fredrich<sup>2</sup>

<sup>1</sup>*mgoffin@student.uliege.be (s191956)*

<sup>2</sup>*thomas.fredrich@student.uliege.be (s191717)*

## I. INTRODUCTION

Interior design visualization remains challenging for non-professionals. Most people struggle to imagine how their spaces would look in different design styles. Traditional visualization methods require either professional design skills or extensive photo editing expertise.

In this paper, we explore a machine learning approach for transforming room interior images between design styles. Our method investigates an enhanced CycleGAN architecture that aims to preserve structural elements while changing aesthetic features.

Our work examines:

- Adapting CycleGAN for room style transfer
- Implementing multi-scale discriminators and improved residual blocks
- Testing a structural preservation loss to maintain architectural elements
- Evaluating results using LPIPS and FID metrics

We focus specifically on transferring between industrial and scandinavian bathroom styles. Our approach works with unpaired datasets, addressing the practical limitation that paired before/after images of the same room in different styles are rarely available.

## II. RELATED WORK

### A. Image-to-Image Translation

Image-to-image translation has seen significant progress in recent years. Pix2Pix [1] introduced a general framework for supervised image translation using conditional GANs, requiring paired training data. CycleGAN [2] extended this approach to the unpaired setting by introducing cycle consistency loss, enabling training without directly corresponding before/after images. This unpaired approach is particularly valuable for interior design applications, where obtaining before/after photos of the same room in different styles is difficult.

### B. Style Transfer

Neural style transfer, explored by Gatys et al. [3], showed the ability to transfer artistic styles to photographs using feature representations from pre-trained networks. Subsequent approaches improved efficiency [4] and versatility [5]. In particular, adaptive instance normalization (AdaIN) [5] and whitening and coloring transforms (WCT) [6] demonstrated better separation of content and style. While these methods work well for artistic style transfer, they often struggle with the structural preservation needs of interior design applications.

### C. Diffusion Models

More recently, diffusion-based generative models have shown exceptional capabilities in image generation and manipulation tasks. Denoising diffusion probabilistic models (DDPM) [7] introduced a new paradigm for high-quality image synthesis. These techniques have evolved into powerful text-to-image models such as DALL-E 2 [8] and latent diffusion models [9].

While these models can produce impressive style transformations with text guidance, they typically require large-scale training data and computational resources. Some recent work has explored using diffusion models specifically for image-to-image translation tasks [10], though often with different training requirements compared to GAN-based approaches.

### D. Interior Design Applications

Image translation and style transfer techniques have been applied to interior design in both academic research and commercial applications. Several existing tools offer interior visualization capabilities. Our work builds upon Fu [11], who demonstrated promising results by enhancing CycleGAN for art style transfer.

We adapt Fu’s architectural modifications—including improved residual blocks, multi-scale discriminators, and specialized loss functions—specifically for room style transfer. While newer diffusion-based approaches show promise, we chose to work with CycleGAN for its established ability to work with unpaired image datasets of modest size and its more accessible computational requirements.

### III. METHODOLOGY

#### A. Dataset

Our project uses a subset of bathroom interior images sourced from the "Interior Design Images and Metadata Dataset from Pinterest" (available on Kaggle as "galinakg/interior-design-images-and-metadata"). The original dataset contained multiple room types (bathroom, bedroom, kitchen, living room) with various design styles, accompanied by CSV metadata files linking images to style categories and attributes.

For our implementation, we made several modifications to the dataset:

- We focused exclusively on bathroom images to enable our model to specialize in a single room type
- We cleaned the dataset to remove:
  - Images containing text/writing
  - Images showing only color palettes
  - Images with single objects rather than complete rooms
- We removed the CSV metadata files and instead organized images directly into style-specific directories

While the original dataset contained five bathroom design styles (boho, industrial, minimalist, modern, and scandinavian), we chose to work with **industrial** and **scandinavian** styles. This decision was based on the clear visual contrast between these styles—industrial bathrooms typically feature raw, utilitarian elements with exposed materials and metal fixtures, while scandinavian bathrooms are characterized by light, airy spaces with natural materials and neutral colors. We reasoned that this strong stylistic difference would make the style transfer effects more pronounced and easier to evaluate.

For training our model, we split each style category into train (70%), validation (15%), and test (15%) sets. This resulted in approximately 105 industrial and 79 scandinavian bathroom images for training, with proportional numbers for validation and testing.

#### B. Model Architecture

Our architecture is based on CycleGAN [2] with several enhancements inspired by Fu [11]. The overall system consists of two generator networks and two discriminator networks, enabling bidirectional translation between industrial and scandinavian bathroom styles. Figure 1 illustrates the high-level architecture of our system.

FIG. 1. Enhanced CycleGAN architecture for bathroom style transfer. The system includes two generators ( $G_A$  transforms Industrial to Scandinavian,  $G_B$  transforms Scandinavian to Industrial) and two discriminators ( $D_A$  and  $D_B$ ). Additional loss functions (cycle consistency and structural preservation) help maintain both the content integrity and the distinctive stylistic elements.

##### 1. Enhanced Generator

Our generator network uses a ResNet-based architecture with improvements for interior design style transfer:

- Initial reflection padding and  $7 \times 7$  convolution to preserve edge information, which is critical for maintaining bathroom structural elements
- Two downsampling layers with stride-2 convolutions and instance normalization
- Nine residual blocks for feature transformation while maintaining spatial information
- Improved upsampling using bilinear interpolation followed by  $3 \times 3$  convolutions to reduce checkerboard artifacts common in transpose convolutions
- Final reflection padding and  $7 \times 7$  convolution with Tanh activation

The dual generator structure allows us to create both Industrial $\rightarrow$ Scandinavian ( $G_A$ ) and Scandinavian $\rightarrow$ Industrial ( $G_B$ ) transformations. This bidirectionality enables cycle consistency training without requiring paired images.

##### 2. Multi-scale Discriminator

To better capture both global structure and local details, we implemented a multi-scale discriminator that operates at two resolutions:

- Full-resolution discriminator for capturing overall image coherence and stylistic patterns
- Half-resolution discriminator for focusing on larger-scale features (like layout and fixtures)
- Each discriminator uses a PatchGAN architecture with spectral normalization for training stability

This multi-scale approach helps the model better distinguish between industrial and scandinavian styles while maintaining structural coherence. For example, the full-resolution discriminator might focus on detailed textures like tile patterns or fixture finishes, while the half-resolution discriminator captures more general properties like color schemes and spatial arrangements.

### 3. Perceptual Network

For our structural preservation loss, we incorporated a perceptual network based on pre-trained VGG19 features. This network extracts features at three different depths, allowing us to capture different levels of structural information:

- Low-level features (edges, textures) from early layers, which help preserve basic elements like walls and fixtures
- Mid-level features (patterns, local structures) from middle layers, capturing bathroom fixtures and their arrangement
- High-level features (spatial arrangements) from deeper layers, representing overall bathroom layout

By weighting these feature levels differently in our loss function (with higher weight on low and mid-level features), we encourage the model to preserve structural elements while still allowing stylistic changes appropriate for the target domain. This is especially important for bathroom style transfer, where we want to maintain the layout and fixtures while changing decorative elements to match either industrial or scandinavian design aesthetics.

### C. Loss Functions

Our training objective combines several loss terms:

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda_{cycle}\mathcal{L}_{cycle} + \lambda_{id}\mathcal{L}_{id} + \lambda_{struct}\mathcal{L}_{struct} \quad (1)$$

#### 1. Adversarial Loss

We use the standard GAN loss with mean squared error (MSE) for stable training:

$$\mathcal{L}_{GAN}(G, D) = \mathbb{E}_y[(D(y) - 1)^2] + \mathbb{E}_x[D(G(x))^2] \quad (2)$$

#### 2. Cycle Consistency Loss

To ensure that the transformed images can be mapped back to their original versions:

$$\mathcal{L}_{cycle} = \mathbb{E}_x[||G_B(G_A(x)) - x||_1] + \mathbb{E}_y[||G_A(G_B(y)) - y||_1] \quad (3)$$

#### 3. Identity Loss

To encourage color preservation when appropriate:

$$\mathcal{L}_{id} = \mathbb{E}_x[||G_B(x) - x||_1] + \mathbb{E}_y[||G_A(y) - y||_1] \quad (4)$$

### 4. Structural Preservation Loss

Our key innovation is a structural preservation loss that helps maintain architectural elements while changing style:

$$\mathcal{L}_{struct} = \sum_{i=1}^3 w_i \cdot ||F_i(G_A(x)) - F_i(x)||_1 + \sum_{i=1}^3 w_i \cdot ||F_i(G_B(y)) - F_i(y)||_1 \quad (5)$$

where  $F_i$  represents the feature maps from the  $i$ -th layer of the perceptual network, and  $w_i$  are weights decreasing with depth to allow style changes while preserving structure.

## IV. IMPLEMENTATION

### A. System Overview

We implemented our system in Python using PyTorch as the deep learning framework. The codebase is organized into several key components:

- **data\_preparation.py**: Dataset loading, preprocessing, and splitting
- **train.py**: Training script with Weights & Biases integration
- **evaluate.py**: Evaluation metrics computation (LPIPS, FID)
- **inference.py**: Style transfer application to new images
- **models/**: Neural network architecture implementation
- **utils.py**: Utility functions for image processing

### B. Training Process

We trained our model using the following hyperparameters:

- Batch size: 4
- Learning rate: 0.0002 with Adam optimizer ( $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$ )
- Training epochs: 100
- Image size: 256×256
- Loss weights:  $\lambda_{cycle} = 10.0$ ,  $\lambda_{id} = 0.5$ ,  $\lambda_{struct} = 10.0$

The training process included several strategies for stability:

- Learning rate decay using a step scheduler (half the rate every 30 epochs)
- Image buffer for discriminator training to reduce oscillation
- Spectral normalization in the discriminator

We used Weights & Biases for experiment tracking, allowing us to monitor loss curves, learning rates, and generated samples throughout training.

### C. Evaluation Metrics

We evaluated our model using two complementary metrics:

- **LPIPS** (Learned Perceptual Image Patch Similarity): Measures perceptual similarity between images, used to evaluate both cycle consistency (how well the original image is reconstructed after a full cycle) and structural preservation (how well structural elements are maintained)
- **FID** (Fréchet Inception Distance): Measures the distance between the feature distributions of real and generated images, providing an assessment of overall generation quality

## V. EXPERIMENTS AND RESULTS

*This section will be populated with actual experimental results once they are available. It will include:*

- *Visual examples of style transfer between different bathroom styles*
- *Quantitative evaluation using LPIPS and FID metrics*

- *Ablation studies examining the contribution of different architectural components*
- *Comparison with baseline methods*

## VI. DISCUSSION

*This section will analyze the results and discuss:*

- *Strengths and limitations of our approach*
- *Analysis of which bathroom styles were most challenging to transfer*
- *The effectiveness of the structural preservation loss*
- *Observed failure cases and potential improvements*

## VII. CONCLUSION AND FUTURE WORK

In this paper, we presented a deep learning approach for bathroom interior design style transfer using an enhanced CycleGAN architecture. Our system can transform bathroom images between five different design styles while preserving structural elements. The key innovations include a multi-scale discriminator architecture and a structural preservation loss function specifically designed for interior spaces.

*This section will be expanded with concrete conclusions based on experimental results.*

Future work could explore several promising directions:

- Extending the approach to other room types beyond bathrooms
- Incorporating semantic segmentation to better preserve specific elements
- Developing controllable style transfer with adjustable intensity
- Exploring real-time implementation for interactive applications

- 
- [1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
  - [2] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
  - [3] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
  - [4] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
  - [5] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.
  - [6] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. In *Advances in neural information processing systems*, pages 386–396, 2017.
  - [7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural*

- Information Processing Systems*, volume 33, pages 6840–6851, 2020.
- [8] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.
  - [9] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.
  - [10] Chitwan Saharia, William Chan, Huiwen Chang, Christopher A Lee, Jonathan Ho, Tim Salimans, David J Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022.
  - [11] Xuhui Fu. Digital image art style transfer algorithm based on cyclegan. *Scientific Programming*, 2022:6075398, 2022.