Photorealistic Image Stylization with Spatial Consistency: Proposal

Khaled Gad 120120024
Anas Awadallah 120210098
Steven Yakoub 120210108

May 24, 2025

Problem Statement

The goal of our project is to implement a photorealistic image stylization system that transfers the style of a reference image to a content image, while preserving the structural integrity and realism of the content. Existing neural style transfer methods often introduce distortions and spatial inconsistencies, which reduce the realism of the final output. Our system aims to address these issues by implementing a method that produces artifact-free, spatially consistent stylizations that remain photorealistic.

The system will take as **input** a content image and a style reference image. The expected **output** is a stylized image that maintains the content structure of the original photo while adopting the visual style of the reference image, with minimal artifacts and consistent stylization across the entire image.

Approach

Our implementation will follow the two-stage pipeline described in the paper: a **stylization step** followed by a **smoothing step**.

- The **stylization step** will use a modified Whitening and Coloring Transform (WCT), applied to deep features extracted using a pre-trained encoder-decoder network. This step performs style transfer by aligning the feature distributions of the content and style images.
- The **smoothing step** will apply a manifold ranking algorithm to improve spatial consistency in the stylized image, particularly within semantically uniform regions. This helps reduce noticeable transitions and artifacts, enhancing photorealism.

Both components have closed-form solutions and can be computed efficiently, making the system suitable for real-time or interactive use.

Experiments and Results

We will validate our implementation using the dataset provided by Luan et al., available at the Deep Photo Style Transfer GitHub. The dataset includes curated content and style image pairs for photorealistic stylization evaluation.

We will divide tasks among group members to implement both stages from scratch, reusing existing encoder-decoder architectures where appropriate. We plan to evaluate the final outputs visually and compare them with results from baseline methods such as standard WCT and neural style transfer. We will also measure performance in terms of runtime and artifact reduction.

A successful outcome will be defined by the ability to reproduce the results reported in the paper: spatially consistent stylizations with fewer artifacts and improved user preference, achieved at a significantly faster processing speed than previous methods. .