Shadow Removal using Diffusion

CV project proposal by Akash G and Sai Phani

Advisor
Shilin Hu (shilhu@cs.stonybrook.edu)

Students
Akash G (gakash@cs.stonybrook.edu). SBU ID: 115099704
Sai Phani Pallapothu (spallapothu@cs.stonybrook.edu). SBU ID: 115073070

Introduction

Shadow removal task aims to remove shadows from an image produced by an occluded light source and restoring the image contents. Previous methods have achieved promising results but still lack in detecting shadow boundaries and distinguishing between different shades of shadows. This is because they focus on solving a physical model of image composition (for example: refer[2]).

Recently, diffusion models have shown surprisingly good results in computer vision tasks, especially in tasks such as the image repainting task (refer[1]). This is because the process of destroying and restoring an image is extremely versatile and can be guided with different requirements and reconstruction methods.

We would like to explore the possibility of training a diffusion model for the shadow removal task. The task is different from tasks like image repainting because it involves recovering the content from the regions affected by shadows as opposed to generating it from scratch.

Goals

1. Setup and train different diffusion models such as guided diffusion, cold diffusion, improved diffusion, stable diffusion, etc on the adjusted-ISTD dataset.
2. Explore different training setups and draw inferences on what works.
3. Add additional loss functions such as Chromatic Consistency loss, Structure Preservation Loss, etc for improved results.

Dataset

We use the Adjusted Image Shadow Triplets Dataset (AISTD) for training our diffusion models. The dataset contains 1870 image triplets of shadow image, shadow mask, and shadow-free image.

Approach
We built and trained our own diffusion models from scratch as this would allow us to have full control over the end-to-end pipeline flow and allow us to study the impact of various hyperparameters on the model's performance.

Architecture
Simple UNet model with 4 upsampling layers and 4 downsampling layers as follows:
down_channels = (64, 128, 256, 512, 1024)
up_channels = (1024, 512, 256, 128, 64)
The number of parameters in our model is roughly **62 million**.

The input to the model is a Shadow-free image and the model loss is computed against the same original shadow-free image.
The diffusion model is conditioned to remove the shadows present in the images. The way this is achieved is by artificially introducing the corresponding shadow image to the denoising/reverse diffusion process.

Training Parameters
Image size: 64x64
Linear beta schedule (describes how much variance/noise is added at each timestep): start=0.0001, end=0.02
Batch size: 128
Optimizer: Adam
Learning rate: 1e-3
Epochs: #todo

Experiments

1. Loss functions
    a. MSE/L2 loss: mean squared difference between the actual image and the predicted image.

    b. L1 loss: mean absolute difference between the actual image and the predicted image.
2. Introducing Chromaticity Consistency Loss: heaps measure how well the chromaticity (i.e., the colors) of the output image is consistent with the chromaticity of the input image. This can improve the quality and realism of the output images.
3. Introducing Structure Preservation Loss to the model with L2 loss: helps measure how well the structural information in the output image is preserved compared to the input image.

Results

We achieved our primary goal of removing shadows from images by training the diffusion model.
#todo

References

1. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., & Van Gool, L. (2022). RePaint: Inpainting using Denoising Diffusion Probabilistic Models. *arXiv*. https://doi.org/10.48550/arXiv.2201.09865
2. Le, H., & Samaras, D. (2020). Physics-based Shadow Image Decomposition for Shadow Removal. *arXiv*. https://doi.org/10.1109/TPAMI.2021.3124934
3. Adjusted ISTD dataset: https://drive.google.com/file/d/1rsCSWrotVnKFUqu9A_Nw9Uf-bJq_ryOv/view?usp=sharing
4. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. *arXiv*. https://doi.org/10.48550/arXiv.2112.10752
5. https://github.com/CompVis/stable-diffusion
6. https://huggingface.co/CompVis/stable-diffusion
7. Jonathan Ho, Ajay Jain, Pieter Abbeel. (2020) Denoising Diffusion Probabilistic Models 2006.11239.pdf (arxiv.org)
8. Croitoru, F., Hondru, V., Ionescu, R. T., & Shah, M. (2022). Diffusion Models in Vision: A Survey. *arXiv*. https://doi.org/10.48550/arXiv.2209.04747
9. Jin, Y., Yang, W., Ye, W., Yuan, Y., & Tan, R. T. (2022). ShadowDiffusion: Diffusion-based Shadow Removal using Classifier-driven Attention and Structure Preservation. *arXiv*. https://doi.org/10.48550/arXiv.2211.08089