Digital Image Processing w25-26 - Project Proposal

- **Student Name:** Yosef Goren
- **ID:** 211515606
- **Email:** yosefgoren@campus.technion.ac.il

Image Colorization: Luminance-Preserving Diffusion

Motivation

Image colorization is a classic image processing problem. Unsurprisingly, various ML models have been successfully applied to it: CNN regression models, GANs, and VAEs.
More recently, this list has come to include Stable Diffusion, which has exhibited SOTA results for image generation.

Therefore, we believe a key question in this domain is:

Q1: How/Should we adapt the Stable Diffusion architecture to fit image colorization?

Research Grounding Point

We consider the paper:

P1: Multimodal Semantic-Aware Automatic Colorization with Diffusion Prior (Wang et al., 2024).

It provides an image colorization solution (at inference time) which takes grayscale images and textual directions as input.
For training, they use a pre-trained image generation diffusion model and fine-tune it with a tweaked architecture.
The tweak introduces luminance conditioning in the reconstruction process by essentially concatenating the grayscale image to the intermediate result before each diffusion step.

However, an important observation from this work is that **luminance is guided, not preserved by construction**. The diffusion process itself still uses latent representations that mix luminance and color, and luminance consistency is enforced indirectly through training. This raises the question:

Q2: What happens if luminance is removed from the reconstruction process altogether?

Project Suggestion

We suggest performing a quantitative comparison between diffusion models that utilize different techniques for using the grayscale image as a prior, aiming to answer Q1.

Additionally, the specific approach where diffusion is applied **only to chrominance channels**, while keeping the luminance channel fixed, will be examined, aiming to answer Q2.

Method Overview

While it is too early to decide on a specific methodology, we imagine answering Q2 might look like:

- Pick one of the following:
  - Taking a pre-trained image generation Stable Diffusion model and adapting it to a form where only luminance is provided to it.
  - Training a model from scratch.
- Training/Fine-tuning it iteratively:
  - Take a colored image.
  - Apply noise in the chrominance space to the image (luminance fixed).
  - Require the model to restore colors.

Whether we use a base model or not, it will be structurally incapable of modifying the luminance.

For inference:

- Take a grayscale image.
- Apply noise in the chrominance space to it.
- Apply the model to it iteratively.

Practical Notes

We have struggled to find a reference paper with an implementation that is actually accessible. This might mean that our work will be based on a different reference to ensure this project maintains a reasonable scale and does not turn into a coding-focused project.