COSE474-2024F: Final Project Proposal
Colorization of Animal Images with Prompts

2022320027 Cho Yunji

1. Introduction

There has been extensive research on various image restoration tasks, including efforts to restore masked images or enhance old photographs. Among these, image colorization has been of particular interest to me. Grayscale images store only luminance information, discarding all RGB data. Restoring the original colors from these images is a highly intriguing challenge.

2. Problem definition & challenges

The goal of my project is to restore image colors guided by a text-based prompt. Although many advanced colorization models have been developed, achieving results nearly identical to the original image, my model aims to stand out by enabling users to specify target colors through prompts. However, achieving generalized performance across all object types is challenging. To address this, I will focus on animal images. Specifically, the task will involve receiving a prompt containing the desired fur color and eye color of the animal, which the model will use to guide the color restoration process. I plan to leverage the CLIP model as a foundation and utilize pretrained models for both text and colorization vision tasks.

3. Related Works

Several models related to colorization exist. HairGAN, for example, specializes in coloring hair and can also modify clothing colors(Xie et al., 2022). Another relevant model is Text2Color, which takes object descriptions as input prompts and uses predefined color palettes associated with the objects to perform grayscale image colorization(Bahng et al., 2018).

4. Datasets

For this project, I will use the Animal Faces-HQ (AFHQ) dataset, which consists of three categories: cat, dog, and wildlife(Choi et al., 2020). I will preprocess the dataset by converting the original images into grayscale to create input data for the colorization task.

5. State-of-the-art methods and baselines

One of the current state-of-the-art models for colorization is the Palette model, known for its impressive performance in generating realistic colors(Saharia et al., 2022). The difference is that the state-of-the-art model focuses solely on vision tasks, whereas my model leverages multimodal processing. I will assess the performance of my model by comparing the output similarity when prompts with colors close to the original are provided. Additionally, I will compare the similarity between the original images and the outputs from both my model and the Palette model to evaluate performance.

6. Schedule & Roles

10/28 – 11/3: Data preprocessing
11/4 – 11/10: Develop prompt preprocessing code and identify suitable colorization models
11/11 – 11/25: Fine-tuning the model
11/26 – 12/2: Performance evaluation

References

Bahng, H., Yoo, S., Cho, W., Park, D. K., Wu, Z., Ma, X., and Choo, J. Coloring with words: Guiding image colorization through text-based palette generation, 2018. URL https://arxiv.org/abs/1804.04128.

Choi, Y., Uh, Y., Yoo, J., and Ha, J.-W. Stargan v2: Diverse image synthesis for multiple domains, 2020. URL https://arxiv.org/abs/1912.01865.

Saharia, C., Chan, W., Chang, H., Lee, C. A., Ho, J., Salimans, T., Fleet, D. J., and Norouzi, M. Palette: Image-to-image diffusion models, 2022. URL https://arxiv.org/abs/2111.05826.

Xie, Y., Huang, Y., Huang, R., Wu, Z., Mao, M., and Li, G. Hairgan: Spatial-aware palette gan for hair color transfer. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 01–06, 2022. doi: 10.1109/ICME52920.2022.9859758.