# Image Colorization

**Group Members**: Kehan Li (kehanli@berkeley.edu), Shujing Hu (rylandhu@berkeley.edu), Mingxiao Wei (mingxiaowei@berkeley.edu), Widyadewi Soedarmadji (wid4soe@berkeley.edu)

##### **Overview**

Our project will focus on fine-tuning a generative model for style-specific image colorization on a dataset consisting of snapshots from *The Simpsons*. The fine-tuned model will color a black-and-white *Simpsons* image in the style of *The Simpsons* cartoons. If the initial coloring result is unsatisfying, we will try to incorporate text prompts and additional multimodal models to aid the coloring process. One addition we will apply to the dataset is a data augmentation of line drawings (via edge detection or brightness filtering). We will compare the model’s performance before and after the augmentation.

##### **Base Generative Model**

We hope to use the Generative Adversarial Networks model from this paper <https://arxiv.org/abs/1803.05400> as our base generative model. This colorization procedure uses a conditional Deep Convolutional Generative Adversarial Network (DCGAN), extends current methods to high-resolution images, and suggests strategies for training efficiency. There is a pre-trained model which we can already download the weights for. The model is also highly customizable, making it an ideal base generative model for our project.

##### **Code Base**

We plan to compare the results of using generative and discriminative models based on the results of fine-tuning/training the following GitHub samples:

Colorful Image Colorization (discriminative): <https://github.com/richzhang/colorization>

Image Colorization with GAN (generative): <https://github.com/ImagingLab/Colorizing-with-GANs/>

The first GitHub repository includes code for real-time user-guided image colorization with learned deep priors and models for automatic colorization. The second one utilizes a deep convolutional generative adversarial network to colorize images. We will focus on fine-tuning the generative model, but we think it would be interesting to compare it to the performance of the discriminative model since the architectures of the two models are significantly different.

##### **Datasets**

We found this dataset on Hugging Face which contains around 800 captioned images of characters from *The Simpsons*. We think this will work well for our desired outcome as we can train the discriminative model and fine-tune the generative model to learn to recolor black-and-white snapshots of *The Simpsons* based on *The Simpsons* color scheme.

<https://huggingface.co/datasets/skiracer/simpsons_blip_captions>

If we find that 800 images are not enough to get good results, we will scrape the web for more images to expand our dataset. If we find text prompts corresponding to the images helpful, we would use BLIPv2 or LLAVA 1.5, which are open-sourced multimodal models, to add captions to the images we find.

##### **Resource Budget**

We looked at the requirements for the main code base we plan to use, and it states that NVIDIA GPU (12G or 24G memory) + CUDA cuDNN would be needed to train the model.