Artistic Image Colorization with Visual Generative Networks

Final report

Yuting Sun ytsun@stanford.edu
Yue Zhang zoezhang@stanford.edu
Qingyang Liu qnliu@stanford.edu

1 Motivation

Visual generative models, such as Generative Adversarial Networks (GANs) [1] and Variational Autoencoders (VAEs) [2], have achieve remarkable results in generating visual images [3, 4, 5, 6]. While most existing work [3, 4] focus on photorealistic images, the problem of generating artistic images is relatively underinvestigated. Different from photorealistic images, artistic images exhibit larger variations in color, visual style and emotion. Therefore, it is challenging for generative models, to capture the richer space of artistic visual domain. In this project, we aim to design visual generative models for the problem of artistic image colorization. We would like to explore multiple settings of colorizing artistic images of different styles. We are interested in the following settings. First, given a gray-scale input image, we expect our system to automatically generate vivid color scheme of the input. Second, given a colorful input image, we would like the generated color scheme to follow user control. To enable this, besides input gray-scale image, our system takes as input one additional $k \times k$ color grid, where user can specify color spatially. Moreover, we would like to evaluate our system on various visual styles/media types, for example oil painting and water color, which are both extremely rich in color. The overall design of our systems is illustrated in Figure 1.

1.1 Prior work

Image colorization

Image colorization has been studied previously. Most existing methods can be categorized as parametric or non-parametric. Non-parametric methods typically transfers color from one image to another [7, 8], while parametric methods often learn a function to predict the missing color [9, 10]. Most related to our work are [11] and [6]. [11] studies the problem of automatically colorize a gray-scale photorealistic image and [6] designs an interactive user controllable system for natural images. Our work significantly differs from the above work that we study image colorization in the domain of art images, which poses further challenges to existing systems due to large variation in color.

Conditional generative models

Conditional visual generative models have been extensively studied recently. Mirza et al. [12] showed that by feeding class labels, GANs can generate MNIST digits. Odena et al. [13] demonstrated such capability on natural images. Besides conditioning on discrete variables, Isola et al. [5] employs a model which transforms an existing images to a desired output image. Sangkloy [6] proposed a system that takes both a structural sketch and color scribbles, so that users can control the high-level structure and color of the synthesized image. Recently Elgammal et al. [14] attempts to generate art images using GANs. Our project differs the previous work in that we would like to introduce color control for the task of synthesizing art images.

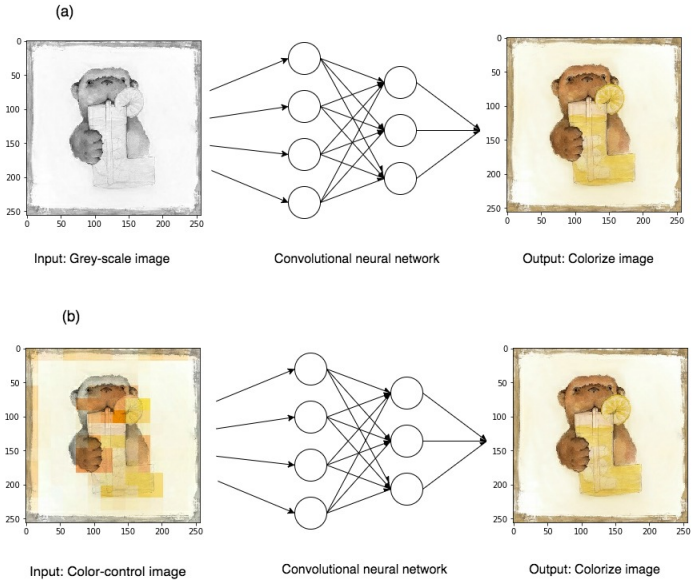

Figure 1: High-level illustration of our systems. In (a), the input is a gray-scale image, which is processed by a convolutional neural network, and becomes a colorized image. In (b), the input is a gray-scale image overlaid with a 14×14 color-controlling grid and the output is a colorized image, which is color-wise consistent with the control grid.

2 Technical methods

In this section, we introduce technical details of our system. We haven tried 2 different neural network models, which will be explained in details in this section.

2.1 Encoder-decoder neural network approach

Our first approach follows the design of the classic encoder-decoder neural networks, where both the encoder and decoder are implemented as Convolutional Neural Networks (CNNs). The encoder network takes as input a gray-scale image (and optionally a $k \times k$ color grid), which is processed through several layers of convolution and pooling operations and becomes a smaller spatial feature map with a larger number of channels. The output feature map is the input to the decoder network, which employs several layers of deconvolution to upsample it to larger feature maps and eventually an output image of the same size as the original input image. The detailed design, (i.e., feature map sizes and number of convolutional filters), is shown in Figure 2. To train this network, we define the loss to be the L2 reconstruction error between the network output and the target color image. In other words, we would like the network to *learn to generate the groundtruth color image with partial input*, which is a gray-scale image with or without coarse color control grid.

2.2 Encoder-decoder neural network with skip link approach

Our second approach also follows the encoder-decoder neural networks design, and we add skip link [15] to this network, therefore in decoder network, each layer will take the corresponding encoder layer as extra input. Skip link is widely used in many computer vision tasks that employ encoder-decoder design. The motivation is to provide the decoder with detailed information directly from the encoder layers for better decoding accuracy. In this work, we employ skip link to further help the decoder to generate HS channels. The detailed design, (i.e., feature map sizes and number of convolutional filters), is shown in Figure 3.

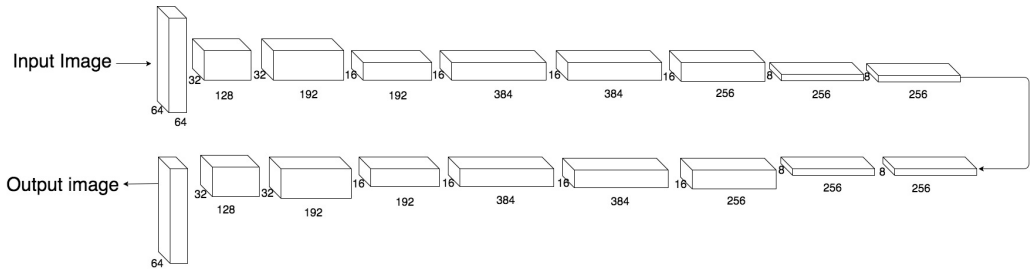

Figure 2: Network design details of our encoder-decoder approach. Due to width limitation, encoder and decoder are displayed in two rows.

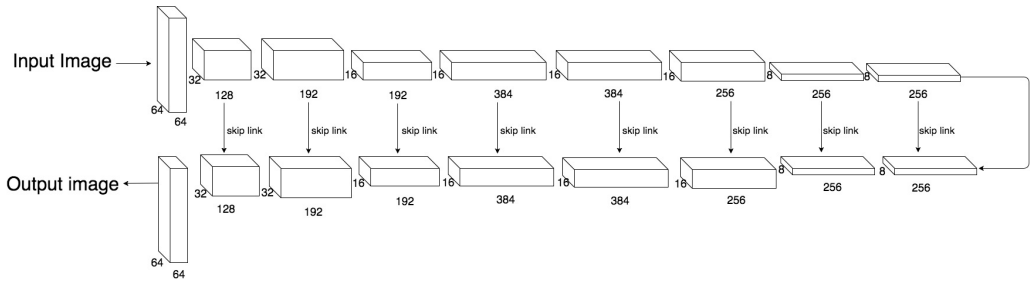

Figure 3: Network design details of our encoder-decoder with skip link approach.

3 Experiments

3.1 Dataset

We use the BAM [16] dataset, which is a recently released dataset of artistic images at the scale of ImageNet [17]. Each image in BAM is labeled with common object types, media types (i.e., visual style) and emotion. The images are labeled iteratively by human annotaters and automatically trained classifiers. The label quality is ensured by the properly designed crowdsourcing pipelines. Specifically, we use images with media type labels in BAM to form our training, validation and test set. We are interested in colorizing two popular media types, which are oil painting and watercolor. (Note that we have experimented with watercolor images in this milestone and plan to evaluate on oil painting in later project phases.)

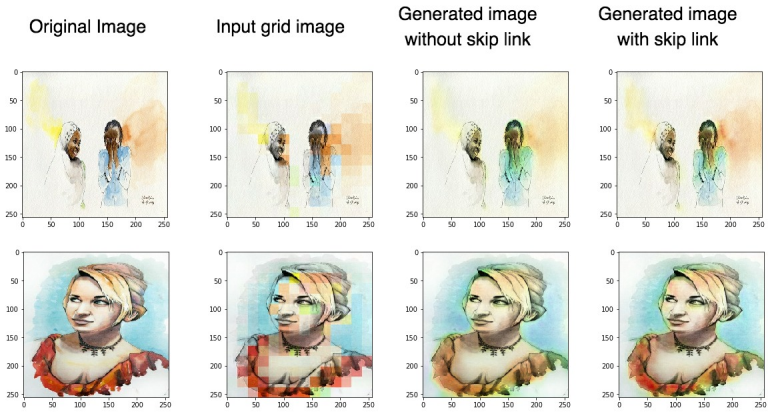

Figure 4: Visual results of colorizing watercolor images with color grid control, from 2 neural networks.

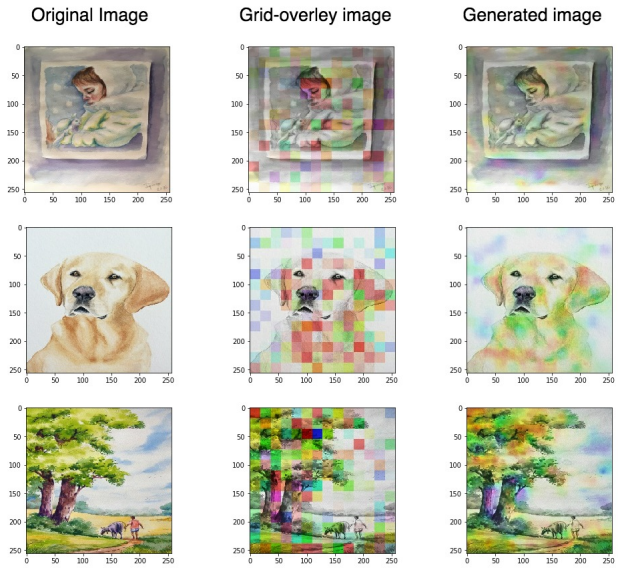

Figure 5: Visual results of colorizing watercolor images with color grid control, from neural networks with skip link.

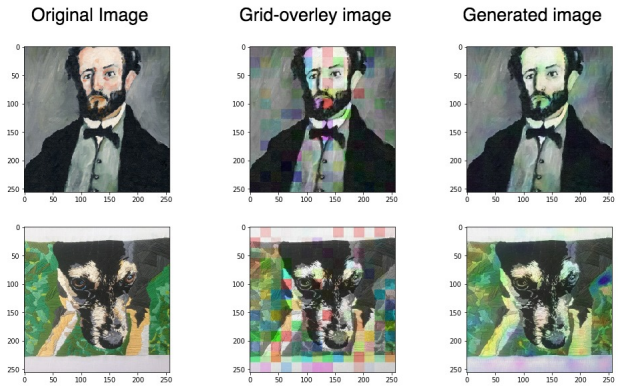

Figure 6: Visual results of colorizing oil painting images with color grid control, from neural networks with skip link.

3.2 Architecture study

We have evaluated both neural networks in the setting of user controllable colorization of watercolor images. We fix the color control grid to have the size of 14×14 and the size of input and out images to be 256×256. To generate training data, for each image, we take its V channel in HSV space (value channel, represent intensity) as input and HS channels (hue and saturation channel, represent color) as groundtruth output. Note that in the training stage of controllable colorization, the color grid is generated by downsampling the HS channels. In test stage, we generate pseudo color control grid by adding random Gaussian noises to the groundtruth 14×14 grid, therefore, we expect our system to synthesize different color than the original image.

As shown in Figure 4, the first column is the original image, the second column is the grey-scale image with color grid overlay, the third column is the image generated by neural network without skip link, and the 4th column is the image generated by neural network with skip link. During architecture study, it shows that our network colorize the grey-scale image under the guidance of color control grid successfully, and neural network with skip link shows more vivid color and clearer boundaries than the network without skip link. Therefore, we use this setting to generate the final results.

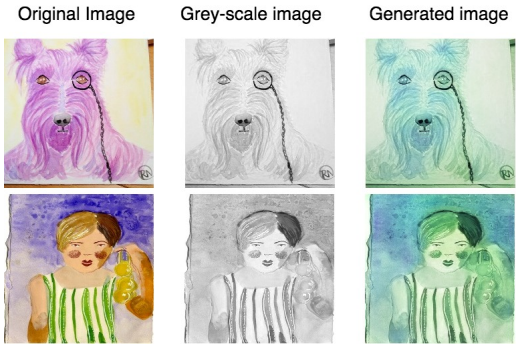

Figure 7: Visual results of colorizing watercolor images without color grid control, from neural networks with skip link.

3.3 Results of neural network with skip link

As shown in Figure 5 for results of colorizing watercolor images, and Figure 6 for results of colorizing oil painting images, our network colorize the grey-scale image under the guidance of color grid and automatically correct incompatible colors in local regions for water color and oil painting images, while we consistently observe that the network fail to capture and reproduce some colors, such as red.

4 Discussions

We have also evaluated neural networks with skip link in colorizing images without color control grid. As shown in Figure 7, the final results is not as visually pleasing as the result of the settings with color control. Specifically, we observe that the generated images tend to have similar color. We would like to point out that colorizing art images without explicit guidance is a much more challenging setting, due to large variation in color in training images. In the future, we plan to further improve our approach by introducing additional losses, e.g., adversarial loss, so that it has the capability to model the large visual appearance variation in the domain of art images.

5 Team member contribution

Yuting Sun: Project definition; Literature study; Data collection; Data processing, Experiment design; Baseline implementation; Architecture study; Error analysis and algorithm tuning; Milestone writeup; Poster writeup; Poster session recording; Final writeup.

Yue Zhang: Project definition; Literature study; Data processing.

Qingyang Liu: Project definition; Literature study; Data collection.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[3] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[4] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015.

[5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2016.

[6] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Scribbler: Controlling deep image synthesis with sketch and color. *arXiv preprint arXiv:1612.00835*, 2016.

[7] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 277–280. ACM, 2002.

[8] Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong. Image colorization using similar images. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 369–378. ACM, 2012.

[9] Aditya Deshpande, Jason Rock, and David Forsyth. Learning large-scale automatic image colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 567–575, 2015.

[10] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415–423, 2015.

[11] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European Conference on Computer Vision*, pages 649–666. Springer, 2016.

[12] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[13] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. *arXiv preprint arXiv:1610.09585*, 2016.

[14] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*, 2017.

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[16] Michael J Wilber, Chen Fang, Hailin Jin, Aaron Hertzmann, John Collomosse, and Serge Belongie. Bam! the behance artistic media dataset for recognition beyond photography. *arXiv preprint arXiv:1704.08614*, 2017.

[17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.