10-701 Project Proposal: MachineLearning4Art

Kenneth Shaw[1], Aditya Sinha[1], Andy Park[1], and Clara Na[1]

[1]Carnegie Mellon University
[1]kshaw2, [2]asinha5, [3]chanhyup, [4]csna@andrew.cmu.edu

1 Proposal

• As a baseline, we will start with the style transfer task on the image dataset from Kaggle [1]. For convenience, we may use an existing implementation [2].

• Possible concrete extension: explore "style" as defined in their artists.csv e.g. "is it easier/harder to transfer between styles of artists from similar backgrounds?"

• Possible additional extension: "Can we control style transfer in a more fine-grained way with a multi-modal system, by incorporating natural language descriptions of desired style into our inputs?" (a la DALL-E [3]).

Evaluation: We will look at how well are image classification predictions preserved between original and stylized outputs, how well do our stylized outputs fit with other images of that style as evaluated by another trained classifier and by human classmate evaluation.

2 Methods

We will use a pretrained model, VGG19 [4] , to blend pairs of images together - with each pair consisting of a *content* image and a *style* image - so the output image has the same features/context but in the style of the *style* image. Content of the image is reconstructed by learning the activation of each kernel in the model. We first start with a white noise image and perform gradient descent to find another image that matches the feature responses of the content image. Style of the image is represented by computing the correlations, which are given by the Gram matrix, between the different filter responses. We also use a white noise image to generate a style that matches the style of a given image. Ultimately, we define the loss function as:

$$\mathcal{L} = \alpha\mathcal{L}_{content}(\vec{p}, \vec{x}, l) + \beta\mathcal{L}_{style}(\vec{a}, \vec{x})$$

where $\mathcal{L}_{content}$ and $\mathcal{L}_{style}$ are the MSE loss function for content and style representation, α and β are the weighting factors for content and style reconstruction, $\vec{p}$ and $\vec{x}$ the original content image and the image that is generated, and $\vec{a}$ the original style image.

3 Dataset

The data we use is a lower resolution version of the dataset downloaded online from `Kaggle` [1]. The dataset consists of over 8000 images from 50 influential artists spanning multiple eras and nationalities, along with non-artwork images from four different categories (architecture, art and culture, food and drinks, and travel and adventure).

References

[1] V. Basu, "Style transfer deep learning algorithm," *Kaggle*, 2019. Available at `https://www.kaggle.com/basu369victor/style-transfer-deep-learning-algorithm/data`.

[2] A. Athalye, "Neural style." `https://github.com/anishathalye/neural-style`, 2015.

[3] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," 2021.

[4] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *CoRR*, vol. abs/1508.06576, 2015.