Global Image Sentiment Transfer

Jie An, Tianlang Chen, Songyang Zhang and Jiebo Luo
Department of Computer Science, University of Rochester, NY, USA
Email: {jan6,tchen45,jluo}@cs.rochester.edu, szhang83@ur.rochester.edu

Problem

Image sentiment transfer focuses on modifying an image from a high-level aspect to change its overall feeling to people. For example, without modifying the content, a family portrait can be transferred to be more positive so that the transferred image may give people a feeling of warmth and peace.

Reference Image Retrieve

We use the Structural Similarity Index (SSIM) [?] to measure the semantic similarity between each image in the target subset and the input image. For every image in the target subset Ω_{tar}, we first compute the SSIM index between the evaluated and the input images and then pick the image with the highest SSIM index as the corresponding reference image to the input image,

$$\text{Reference} = \max_{a \in \Omega_{tar}} \text{SSIM}\left(a, \text{Input}\right). \quad (1)$$

Global Sentiment Transfer

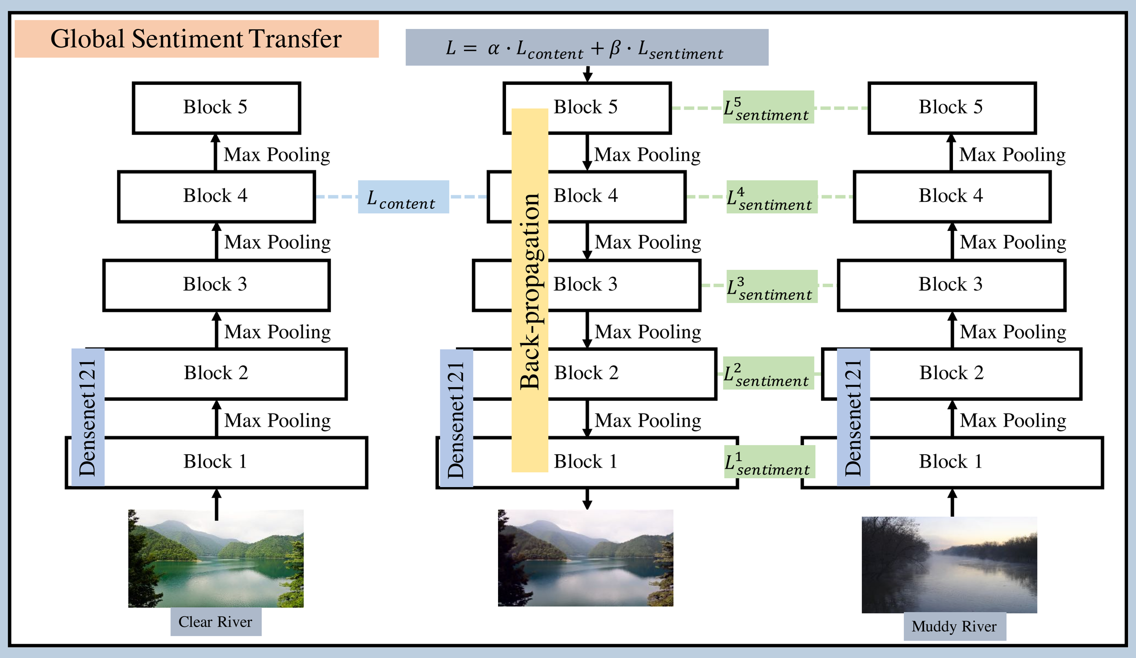
Our algorithm is based on an optimization method, which iteratively transfers the sentiment of images by minimizing two objectives on deep features. The first objective is to keep the details of the input intact, while the second is to force the sentiment of the produced image similar to the reference image. The overall loss functions we use is,

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{content} + \beta \cdot \mathcal{L}_{sentiment}, \quad (2)$$

$$\mathcal{L}_{content} = \|f^4 - f_s^4\|_2, \quad (3)$$

$$\mathcal{L}_{sentiment} = \frac{1}{5} \cdot \sum_{i=1}^{5} \|\text{Gram}\left(f^i\right) - \text{Gram}\left(f_t^i\right)\|_2, \quad (4)$$

where $\text{Gram}\left(f\right) = f^T \cdot f$, f^i denotes the feature maps of the transferred images in DenseNet121. All the feature maps f has a shape of $C \times (H \times W)$, where C denotes the channel number while H, W represent the height and width of f, respectively.
The proposed algorithm adopts the DenseNet121 network pre-trained on the ImageNet dataset [?] as the feature extractor.

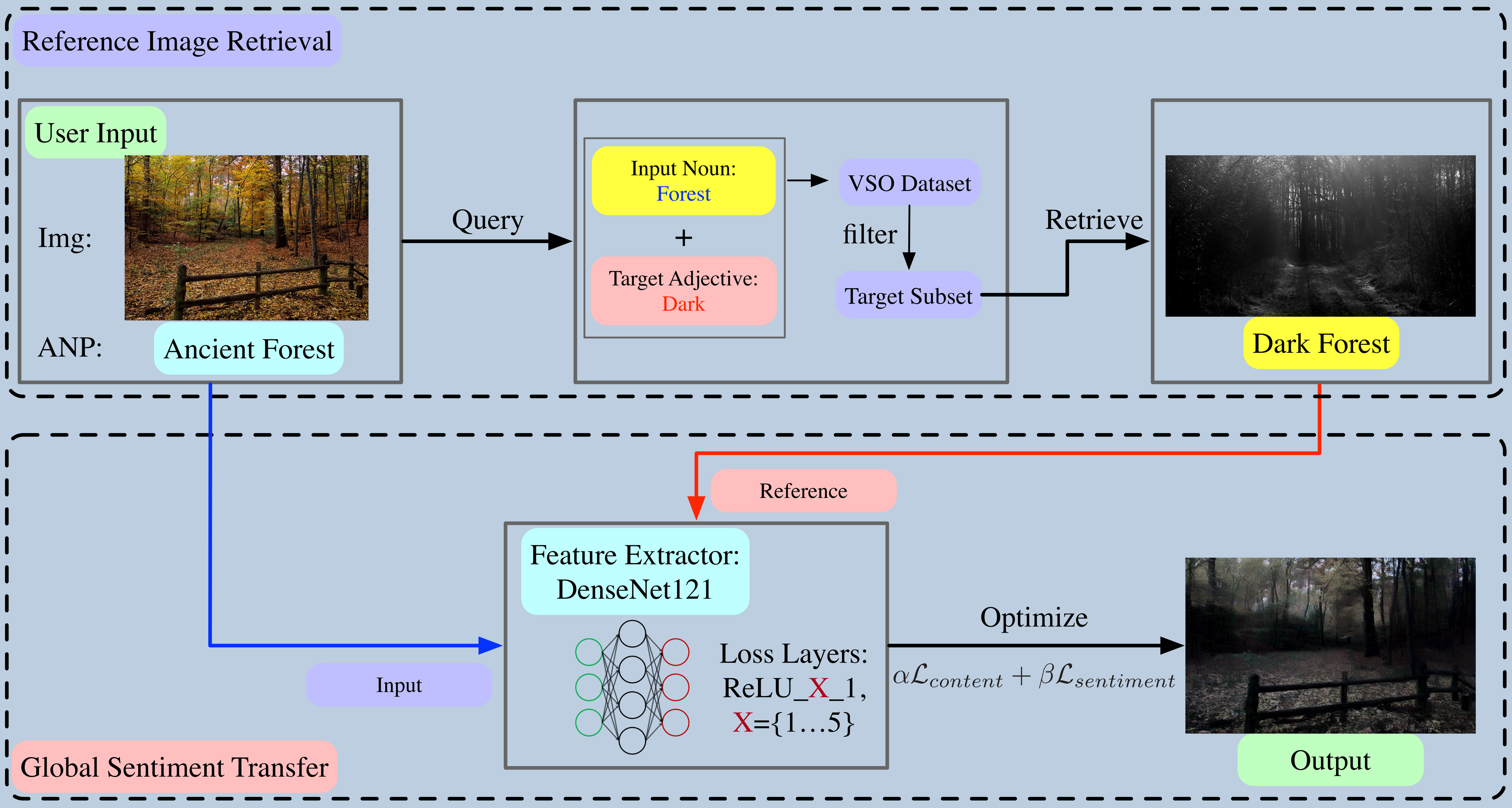


Acknowledges

This work is supported in part by NSF awards IIS-1704337, IIS-1722847, and IIS-1813709, as well as our corporate sponsors.

Framework

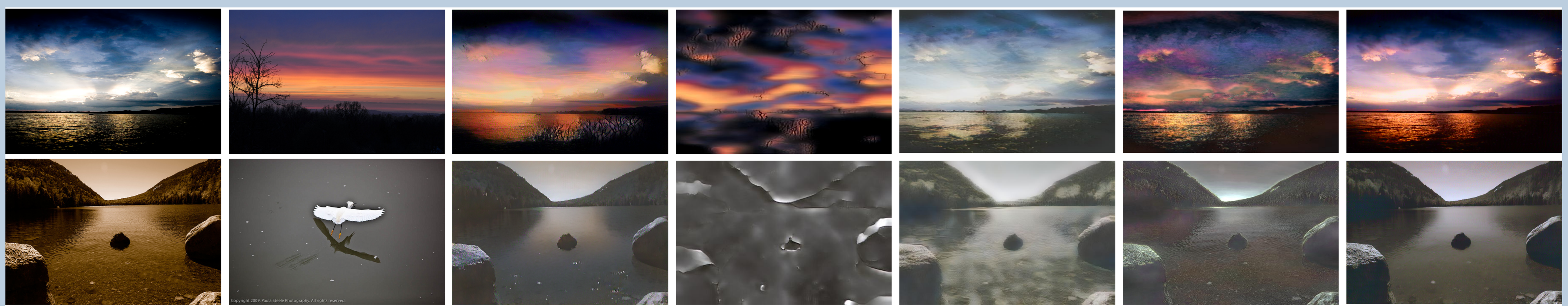
We propose a method that consists of a *reference image retrieve* step and a *global sentiment transfer* step. Given an input image, we first retrieve a reference image with the same content description but an opposite sentiment to the input image. Then a global sentiment transfer algorithm is employed to transfer the input image sentiment to the reference.



Dataset

We first filter the mislabeled images using the list given by [?]. We then manually pick the images that have a global sentiment. For example, an image whose Adjective-Noun Pair (ANP) is "beautiful bird" should not be selected since "bird" is only an image region. On the contrary, some ANPs such as "clear water" and the "lovely city" are selected because they describe global scene properties. The dataset we use contains 10,532 images with positive sentiments and 5,477 images with negative sentiments, which belong to 86 positive ANPs and 43 negative ANPs, respectively.

Experiments


(a) Input (b) Reference (c) Gatys *et al.* (d) WCT (e) AdaIN (f) StyleNAS (g) Ours

Method	Gatys *etal.*	WCT	AdaIN	StyleNAS	Ours
SSIM↑	0.7019	0.2443	0.5301	0.6653	**0.8719**
FID Score↓	169.73	245.21	206.51	191.14	**154.73**

We compare the result produced by our algorithm with both the state-of-the-art artistic [?, ?, ?] and photo-realistic [?] style transfer algorithms. The results by the artistic style transfer algorithms (*e.g.* StyleSwap[?], WCT [?], AdaIN [?]) often have distorted content details, which may be necessary to create artistic effects but are not desirable for producing good sentiment transfer results. The photo-realistic style transfer algorithm [?] can preserve the content information. However, it can create significant artifacts.
We also compute the SSIM index between the *input image* and the produced result to evaluate the content preservation ability of the algorithms. Moreover, we use the FID score between the *reference image* and the produced result to evaluate the sentiment transfer performance. In comparison, we collect 46 input-reference image pairs from the filtered global VSO dataset as the validation set. We obtain the sentiment/style transfer results by running all the compared algorithms on the validation set. The proposed global sentiment transfer algorithm has a higher mean SSIM score than other style transfer methods, demonstrating that our method has a stronger ability to preserve the details of the input image. In addition, the proposed algorithm also achieves a lower FID score, which indicates that the proposed algorithm outperforms other algorithms in creating a more similar sentiment to the reference image.