2020 25th International Conference on Pattern Recognition (ICPR)
Milan, Italy, Jan 10-15, 2021

Global Image Sentiment Transfer

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

***Abstract*—Transferring the sentiment of an image is an unexplored research topic in computer vision. This work proposes a novel framework consisting of a *reference image retrieval* step and a *global sentiment transfer* step to transfer image sentiment according to a given sentiment tag. The proposed image retrieval algorithm is based on the SSIM index. The retrieved reference images by the proposed algorithm are more content-related than the algorithm based on the perceptual loss. Therefore, it can lead to a better image sentiment transfer result. In addition, we propose a global sentiment transfer step, which employs an optimization algorithm to iteratively transfer image sentiment based on the feature maps produced by the DenseNet121 architecture. The proposed sentiment transfer algorithm can transfer image sentiment while keeping the content of the input image intact. Both qualitative and quantitative evaluations demonstrate that the proposed sentiment transfer framework outperforms existing artistic and photo-realistic style transfer algorithms in producing satisfactory sentiment transfer results with fine and exact details.**

I. Introduction

Transferring the sentiment of an image is an unexplored research topic. In contrast to the existing well-known tasks such as two-domain image-to-image translation [1]–[4] (*e.g.* winter → summer, cat → dog) and image style transfer (e.g. artistic style transfer, photo-realistic style transfer), image sentiment transfer focuses on modifying an image from a higher-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. As we live in an age of stress, we believe that this research topic is useful with its strong potential to improve people's lives.

Intuitively, image sentiment is an abstract concept. Compared with two-domain image-to-image translation that commonly has a definite pattern to transfer between two domains (*e.g.* cat → dog, horse → zebra), there are many ways to transfer an image to a specific sentiment. To make image sentiment transfer controllable, a reference image should be fed into the model as the guidance. Considering its similarity to the image style transfer task, we can leverage the existing image style transfer models to perform reference-guided image sentiment transfer. However, it is nontrivial to implement this design because of the poor compatibility between the input image and the reference image for image sentiment transfer. Moreover, directly using existing artistic and photo-realistic style transfer models generally fails to create visually pleasing results in terms of detail preservation and artifact/distortion

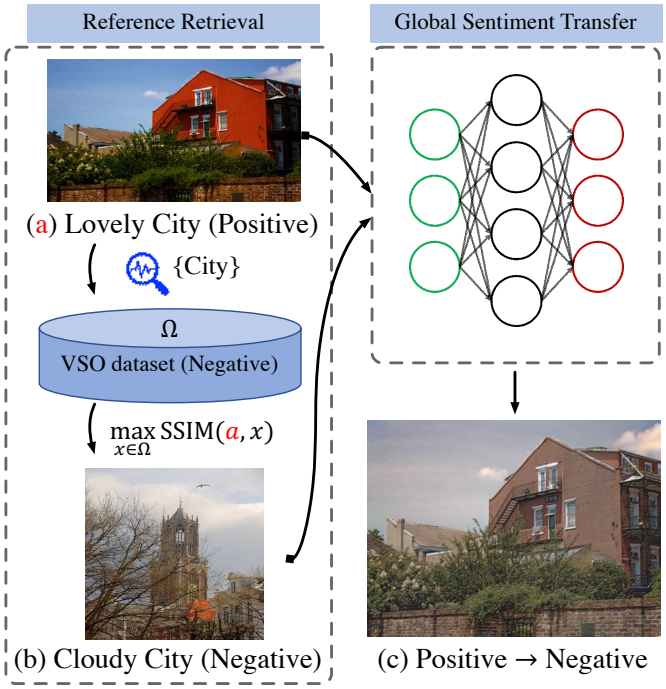


Fig. 1. **Global sentiment transfer.** Given an input image (a) and the adjective-noun tag "lovely city"), we first retrieve a reference image (b) with the same content (noun) but the opposite sentiment ("cloudy city"). We then use a global sentiment transfer algorithm to generate (c), which contains the content of (a) and the sentiment description of (b).

minimization. Compared with image style transfer for which an artistic or photo-realistic style can be indiscriminately added to any input image, the sentiment transfer between two content-unrelated images is risky. For the example of Fig. 2, the sentiment transfer result lacks photo-realism because the reference image does not bring any content-related information.

To this end, we propose a new image sentiment transfer framework that starts with image retrieval. As shown in Fig. 1, given an input image and a sentiment tag provided by the user, instead of randomly sampling a reference image that contains the input sentiment tag, we retrieve the most suitable reference image(s) based on the structural information of the input image. Leveraging the structural similarity score (SSIM) [5], the framework effectively constructs the content relation between the input and the reference image. In Section IV, we demonstrate that this image retrieval step is crucial to improving the performance of image sentiment transfer.

978-1-7281-8808-9/20/$31.00 ©2020 IEEE

2020 25th International Conference on Pattern Recognition (ICPR) | 978-1-7281-8808-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPR48806.2021.9413221

(a) Input (b) Reference (c) Sentiment Transfer Result

Fig. 2. **A failed sentiment transfer case with a content-unrelated image as the reference image.** The generated image is not photo-realistic.

To transfer the sentiment of the input image to that of the retrieved reference image, we design a novel global image sentiment transfer algorithm. Inspired by the image style transfer algorithm by Gatys *et al*. [6], we use an optimization algorithm on deep features from the neural networks pre-trained on the ImageNet dataset [7] to iteratively transfer the sentiment of the reference image to the input image. Unlike existing style transfer algorithms, our method adopts the DenseNet121 architecture as the feature extractor instead of the widely-used VGG19 architecture. We empirically find that DenseNet121 architecture outperforms VGG19 and other popular architectures in fine detail preservation and artifact/distortion minimization. Therefore, DenseNet121 is more suitable to implement sentiment transfer where the produced image should be photo-realistic.

Our main contributions are summarized as follows:

- We are the first to explore the task of image sentiment transfer. We present an effective two-step framework consisting of image retrieval and reference-guided image sentiment transfer.
- We introduce a reference image retrieval algorithm based on the SSIM index, which can achieve better results than other methods in finding content-related reference images.
- We propose a global sentiment transfer algorithm based on DenseNet121, which can effectively transfer the sentiment/style of an image while preserving its fine details.

II. Related Work

Visual sentiment understanding has been explored in recent years. Most existing works focus on the visual sentiment classification tasks. To perform accurate classification for images with different sentiments, low-level features, like color [8]–[10], texture [10], and shape [11] have been studied in the early years. Later on, mid-level composition [10], sentributes [12], principles-of-art features [13], high-level adjective-noun pairs (ANP) [14] are also proposed. Most recently, due to the rapid development of the convolution neural network (CNN) for extracting visual features, many approaches turn to CNN-based sentiment recognition. Some of them make use of noisy data during the training process [15]–[17], while others explore visual sentiment at image region level [18]–[22]. However, compared with image sentiment classification, other sentiment-related tasks such as image sentiment generation/translation have not been studied yet.

The most related tasks to ours are image-to-image translation and image style transfer. Image-to-image translation targets at learning an image-to-image mapping from two different domains. Early approaches need paired data to train the model and are essentially restricted to learning a deterministic one-to-one mapping [23]–[25]. This prevents the generation of diverse output images. CycleGAN [2] first proposes a cycle consistency loss to enable the model to be trained from unpaired data. The following approaches such as MUNIT [1] and DRIT [3] further propose disentangled representations that enable the output images to be diverse. On the other hand, our task is also related to image style transfer. A great number of approaches are proposed for artistic style transfer [26]–[30] and photo-realistic style transfer [31]–[35]. Unlike the above approaches, we focus on image sentiment transfer that requires a strong content relation between the input and the reference image. Therefore, we retrieve reference image(s) based on the sentiment tag provided by the users instead of directly asking users to provide the information. The proposed global sentiment transfer algorithm is based on the work by Gatys *etal*. [28]. However, our algorithm uses the DenseNet121 [36] network architecture instead of VGG19 [37] as the feature extractor since we empirically find that DenseNet121 can achieve more faithful input detail preservation compared with VGG19.

Retrieving a reference image by a given sentiment tag is related to the image retrieval task. Image retrieval aims at finding an image that is close to the given image. Most recent works measure the perceptual loss of images by comparing the image features extracted from pre-trained convolution neural networks [38]. Unlike these works based on the perceptual loss, we find that the SSIM index [5] is more suitable for retrieving reference images for sentiment transfer since it captures more structural similarity of images.

III. Method

To transfer the global sentiment of an image, we propose a method that consists of a *reference image retrieve* step and a *global sentiment transfer* step. Fig. 3 shows the framework of the proposed algorithm. 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. We describe the details of these two steps in the following. .

A. Reference Image Retrieve

Retrieving a reference image is the initial step to make sentiment transfer. To achieve high quality sentiment transfer result while facilitating the following global sentiment transfer step, the retrieved image should have a similar content to the input image but contain the feelings of the target/opposite sentiment. To achieve this, we propose an image retrieval algorithm based on the Visual Semantic Odometry (VSO) [39] dataset. For each image in the VSO dataset, an adjective-noun pair is attached to describe the semantic content (noun) and its sentiment (adjective), respectively. To retrieve a reference image according to a given sentiment tag, we first select a

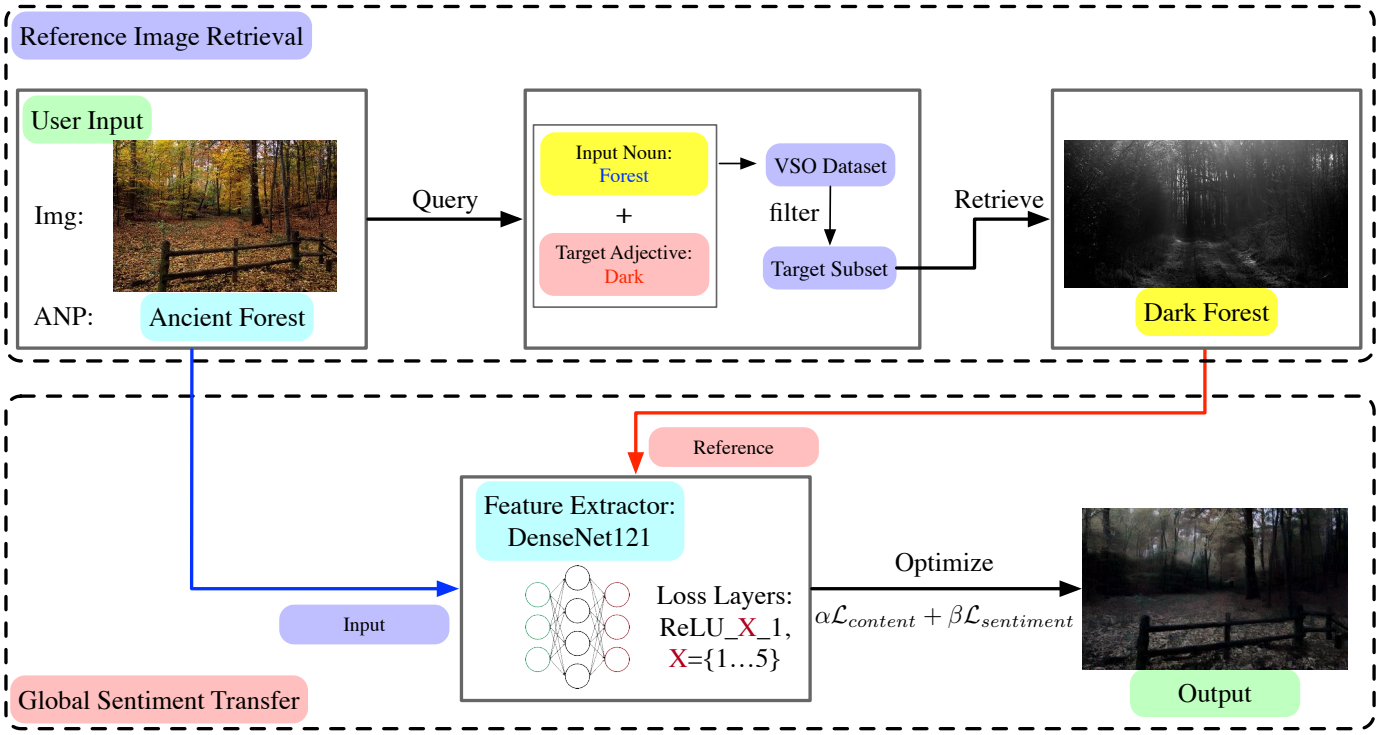


Fig. 3. **Framework of the proposed algorithm.** Our method consists of two parts: a reference image retrieve algorithm and a global sentiment transfer algorithm based on the retrieved reference image.

subset of the VSO dataset, where every image within the subset contains the content tag (noun) of the input image but the sentiment tag (adjective) of the given target, For example, in Fig. 3, the input image has an adjective-noun pair of "Ancient Forest" while images in the corresponding subset have a label of "Dark Forest". If the target sentiment tag is not available, the retrieval algorithm finds the reference image directly from a larger dataset whose images have the same content description (*e.g.*, city) as the input image but contain the adjectives with the opposite sentiment. Although the input and reference images have the same "noun" tag, and therefore the reference image is likely to have similar content to the input image, the proposed reference retrieval algorithm is nontrivial because the VSO dataset is very noisy and contains many mislabelled data even after the cleaning. Therefore, it is necessary to select reference images carefully.

To find a reference image from the above-mentioned target subset, a metric to measure the distance between images is necessary. Inspired by [33], [35], we use the Structural Similarity Index (SSIM) [40] to measure the semantic similarity between each image in the target subset and the input image. The SSIM index is originally used by image/video quality assessment methods. We empirically find that SSIM is more suitable than the widely-used perceptual loss to measure the semantic similarity between two images. 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,

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

Fig. 6 (b, e) shows the image retrieval results by the proposed algorithm. The retrieved images have the same content as the input images but different sentiment.

B. Global Sentiment Transfer

With the given input and a selected reference image structurally similar to the input image, we propose a novel algorithm to transfer the input image sentiment according to the selected reference image. 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.

A high-quality sentiment transfer result should have a similar sentiment to the reference image while keeping the content details intact compared with the input image. The key challenge in sentiment transfer is to measure the sentiment similarity between two images. Inspired by Gatys et al. [6], [41], we adopt the Gram loss on the deep features of the input and reference images produced by neural networks to measure the sentiment similarity. Such a Gram-based loss term is originally used to measure image style similarity. Since sentiment can be regarded as higher abstraction of the style, we borrow the Gram loss term to implement sentiment transfer.

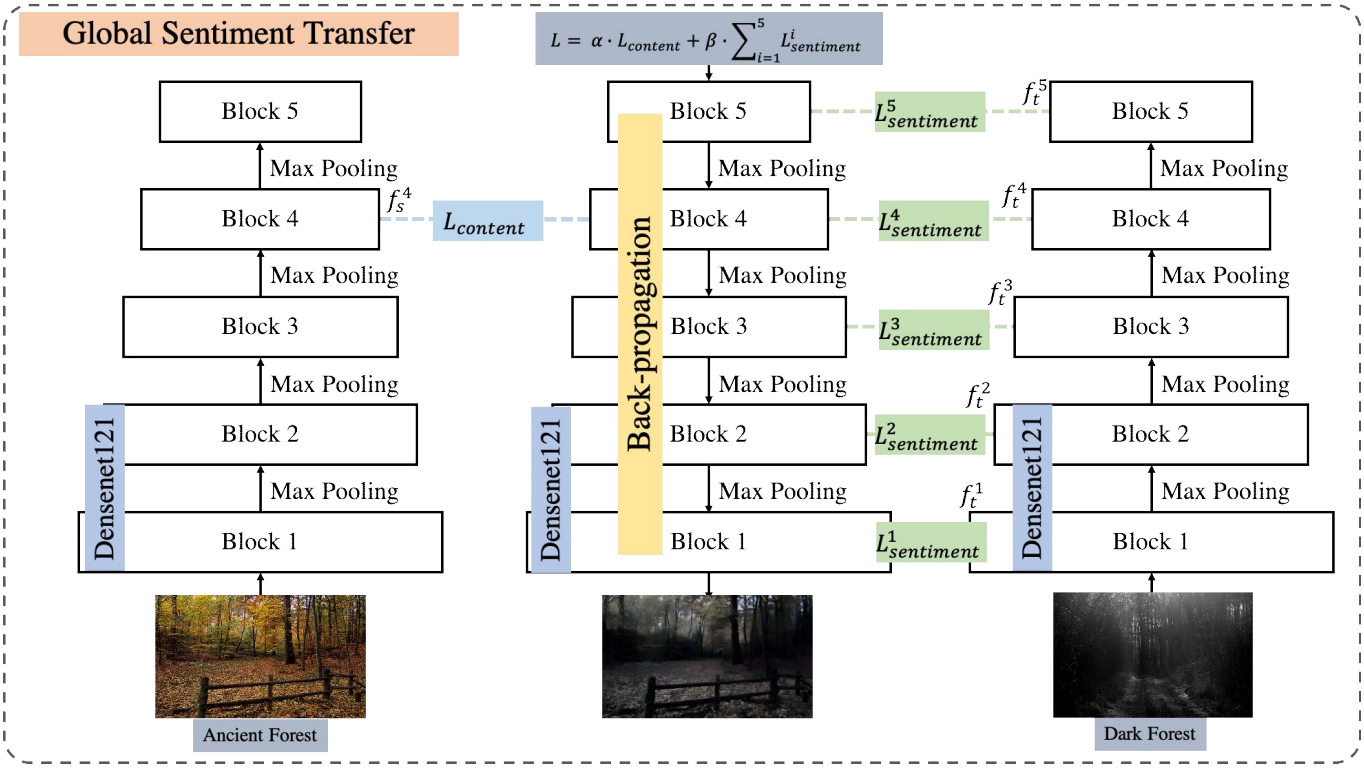


Fig. 4. **Illustration of the proposed global sentiment transfer algorithm.** Here we use the DenseNet121 architecture as the backbone network.

Moreover, we compute the l_2 norm between the transferred and input image features as the content-consistency loss.

The sentiment/style transfer results created by the loss terms mentioned above rely heavily on deep features used to compute the objective functions. Many style transfer algorithms [32], [41]–[46] use the features produced by the VGG19 network pre-trained on the ImageNet dataset [7]. However, the optimization algorithm based on the features of VGG19 can inevitably change the details of the content image. Take the style transfer results of Gatys et al. [42] shown in Fig. 5 for example. The style transfer algorithm based on the features by VGG19 changes the details of the sea, sky, and plants in the input image.

As illustrated in Fig. 4, the proposed algorithm adopts the DenseNet121 network pre-trained on the ImageNet dataset [7] as the feature extractor. We empirically find that DenseNet121 can achieve good sentiment transfer effects while minimizing the influence to the input content's details. Using DenseNet121, we first obtain the deep features of the input and the reference images. These features are produced by the ReLU layers behind five pooling operators in the network. We use $f_s^i,\ i \in \{1...5\}$ and $f_t^i,\ i \in \{1...5\}$ to denote the feature maps of the input and reference image, respectively, where s denotes the *source* while t represents the *target*. DenseNet121 has two advantages: first, DenseNet121 can minimize the loss of the content information while achieving high-quality sentiment transfer. Second, DenseNet121 is more time-efficient because it contains only half of VGG19's parameters (DenseNet121: 6.952 $v.s.$ VGG19: 12.945).

The overall loss functions we use is,

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

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

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

where $\mathrm{Gram}\,(f) = f^T \cdot f$, 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. Note that the transferred image is modified iteratively as the variable in the optimization process to force its content to be the same as that of the input image while having the sentiment of the reference image.

IV. Experiment

In this section, we first discuss the experimental settings and the dataset we use. We then compare the proposed image retrieval and global sentiment transfer algorithms with other image retrieval and image style transfer algorithms, respectively. We finally demonstrate the effectiveness of the proposed algorithms by both visual and quantitative evaluations.

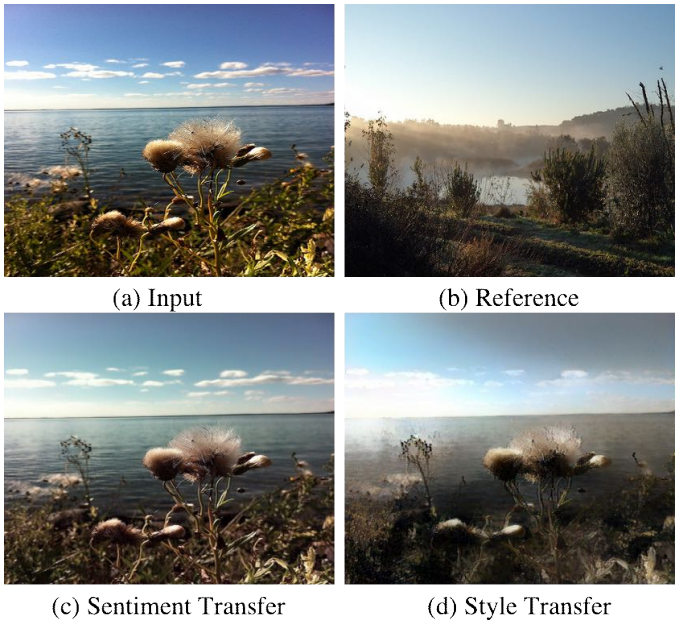


Fig. 5. **Comparison between style transfer results and sentiment transfer results.** The style transfer results are produced by the VGG19 architecture while the sentiment transfer images are generated based on DenseNet121.

Fig. 6. **Comparison between the proposed image retrieval method based on the SSIM index and the image retrieval method based on the perceptual loss.** The proposed retrieval method focuses more on finding content-related images than the algorithm based on the perceptual loss.

A. Experimental Settings

Global Sentiment Dataset. To demonstrate the effectiveness of the proposed global sentiment transfer framework, we collect a few images from the VSO dataset. Since the original VSO dataset is noisy, we first filter the mislabeled images using the list given by [47]. 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.

TABLE I
FID SCORES OF THE REFERENCE IMAGE RETRIEVAL BASED ON THE SSIM INDEX AND THE PERCEPTUAL LOSS. A SMALLER FID SCORE INDICATES BETTER IMAGE RETRIEVAL PERFORMANCE.

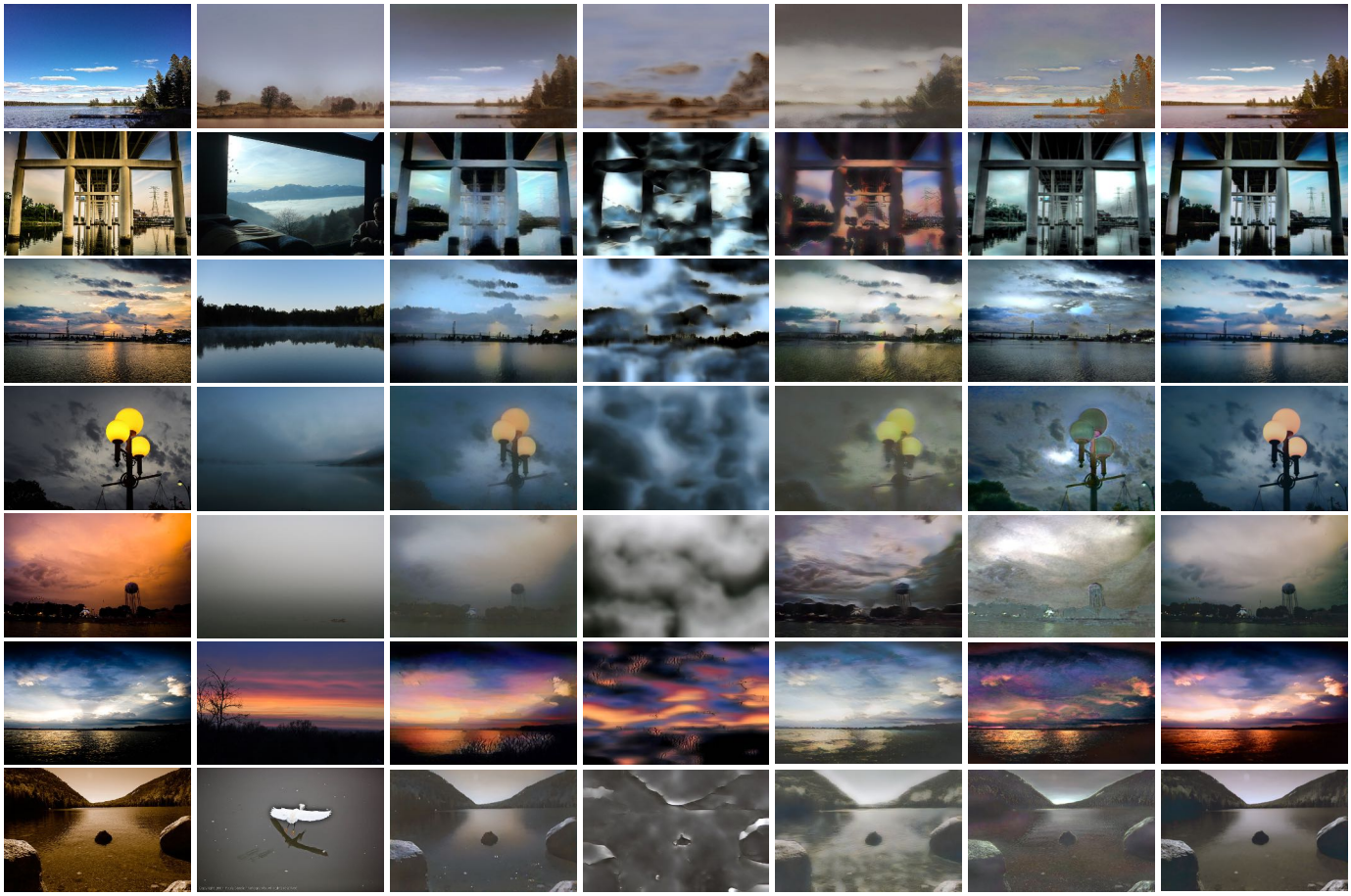
Retrieval Method	SSIM Score	Perceptual Loss
FID Score↓	**158.54**	205.01

Global Sentiment Transfer Settings. To transfer the sentiment from the selected reference image to the input image, we propose an optimization-based iterative method. As stated above, we use the DenseNet121 pre-trained on the ImageNet dataset as the feature extractor. In the optimization process, we use the Adam algorithm [48] to minimize the objective function between the input image and the retrieved reference image. To balance the content and sentiment loss functions, we set $\alpha = 1$ and $\beta = 1,000,000$ in Equation 2. Every input-reference pair takes 500 iterations to produce the sentiment transfer result.

B. Reference Image Retrieval

Fig. 6 shows the image retrieval results based on the perceptual loss and SSIM index. Since the perceptual loss concentrates less on the structure of the image content than the SSIM score, the retrieved images based on the perceptual loss generally have different content structures. On the contrary, the retrieved images by the algorithm based on the SSIM score usually contain the same content information as the input images. Take Fig. 6 for example. The retrieved images by the perceptual loss (c, f) have different content structures from (a) and (d), respectively. On the contrary, the input images (a, d) and the retrieved images by the SSIM index (b, e) have similar content structures. Generally, the global sentiment transfer algorithm would generate a better result if the reference image has a more similar content structure to the input image. Therefore, Fig. 6 demonstrates that the proposed image retrieval method based on the SSIM index outperforms algorithms based on the perceptual loss.

To quantitatively validate the above-mentioned observations, we compare the FID score [50] between the retrieved image set and the input image set. We randomly select 380 images from the filtered global VSO dataset as the validation set to compute the FID score. The retrieved dataset consists of the reference images picked by the image retrieval method based on the evaluation set. FID is originally used to evaluate the performance of image generation methods. Since FID has an excellent ability to measure the distance between two image distributions, we borrow it from GANs [50]–[55] to evaluate the image retrieval performance. Table I shows the FID score of the image retrieval method based on the SSIM index and the perceptual loss, respectively. Table I demonstrates that the proposed method based on the SSIM score outperforms the image retrieval method based on the perceptual loss in finding the most structurally similar reference images.



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

Fig. 7. **Visual comparison between the results produced by the proposed global sentiment transfer algorithm and the state-of-the-art universal style transfer algorithms.** All the compared results are produced by running the officially-released code of the corresponding algorithms.

TABLE II
COMPARISON OF THE MEAN SSIM SCORE AND FID SCORE ON THE VALIDATION SET. A HIGHER SSIM SCORE INDICATES A BETTER DETAIL PRESERVATION ABILITY WHILE A SMALLER FID SCORE INDICATES A BETTER SENTIMENT TRANSFER ABILITY.

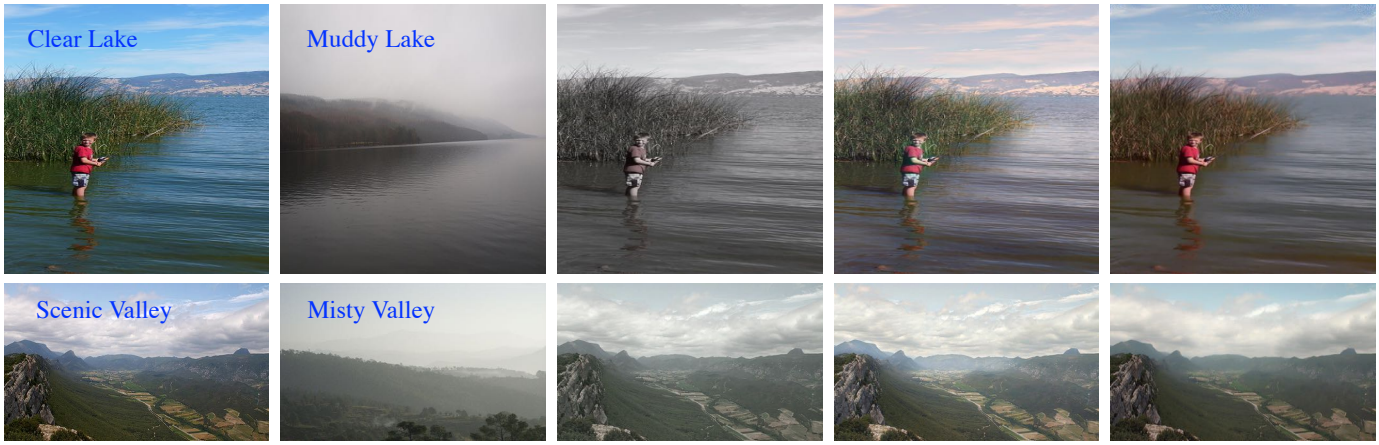
Method	Gatys *etal*. [6]	WCT [49]	AdaIN [27]	StyleNAS [35]	Ours
SSIM↑	0.7019	0.2443	0.5301	0.6653	**0.8719**
FID Score↓	169.73	245.21	206.51	191.14	**154.73**

C. Visual Comparison

Since this work is the first global sentiment transfer algorithm for arbitrary input images, we compare the result produced by our algorithm with both the state-of-the-art artistic [27], [49], [56] and photo-realistic [35] style transfer algorithms to demonstrate the effectiveness of the proposed global sentiment transfer method. Other photo-realistic style transfer algorithms such as [31]–[33] are not compared because these methods need a segmentation map or post-processing to implement style transfer. Ours and the compared methods do not need such a pre or post-processing step. To make a fair comparison, all of the compared style transfer algorithms use the same retrieved images produced by our image retrieval algorithm as the reference. Fig. 7 shows the sentiment transfer results of our method and the style transfer results by the state-of-the-art universal style transfer algorithms. The results by the artistic style transfer algorithms (*e.g*. StyleSwap [56], WCT [27], AdaIN [49]) 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 [35] can preserve the content information. However, it can create significant artifacts. Take Fig. 7 (f) for example. Fig. 7 (g) shows the results produced by our global sentiment transfer algorithm, which achieves the high-quality sentiment transfer effects while ensuring that the content details are not altered. Moreover, the results by our method contains significantly fewer artifacts than the state-of-the-art photo-realistic style transfer algorithms [33], [35].

D. Quantitative Comparison

To quantitatively demonstrate the effectiveness of the proposed algorithm, we adopt the SSIM score and the FID score to evaluate the performance of the proposed algorithm. We



(a) Input (b) Reference (c) RCT (d) SOT (e) Ours

Fig. 8. **Sentiment transfer result comparison between the proposed algorithm and the color transfer methods.**

(a) Input (b) Reference (c) Result

Fig. 9. **Failure cases of the proposed algorithm.**

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. Table. II shows the mean SSIM/FID scores of 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.

E. Comparison with Color Transfer Algorithms

Given an input image and a reference image, traditional color transfer algorithms [57], [58] may be directly used to transfer the sentiment from the reference image to the input image. Fig. 8 compares the sentiment transfer results of our method with the traditional RCT [57] and SOT [58] color transfer algorithms. We empirically find that the color transfer algorithms have the advantage of generating clear results robustly. However, in certain cases, such an advantage can be a limitation. Take the bottom row of Fig. 8 for example, the results of the color transfer algorithms are compromised in transferring the "misty" sentiment, where the generated images should be blurry to create the "misty" feel. Because CNNs can recognize and separate different semantic contents, the proposed method can handle more complex color matching cases and transfer sentiment semantically. Take the top row of Fig. 8 for example, the transferred results by the proposed global sentiment transfer algorithm can change the sentiment of the water from "clear" to "muddy" while keeping the blue sky and red clothes almost intact. However, the color transfer algorithms [57], [58] changes the blue sky and the red clothes because they can only match colors and thus cannot transfer sentiments according to image semantics.

F. Failure Cases

Since global image sentiment transfer is an unexplored topic, the proposed algorithm is far from perfect. On the one hand, we find that the proposed algorithm is not robust on certain input and reference images (*e.g.*, extremely dark images). On the other hand, as shown in Fig. 9, the produced results may be less photo-realistic if the reference and the input image are not matched appropriately. We will address the above-mentioned failure cases in the future. Moreover, since the VSO dataset is fairly noisy and contains many mislabeled samples, it is valuable collect a more clean and specific dataset for the image sentiment transfer task.

V. Conclusion

In this paper, we present a highly promising global image sentiment transfer framework consisting of a reference image retrieval step and a global sentiment transfer step. In the reference image retrieval step, we adopt the SSIM index instead of the perceptual loss to measure the scene structural distance between images. In the global sentiment transfer step, we use the DenseNet121 network pre-trained on ImageNet as the feature extractor and employ an image style transfer framework to iteratively transfer sentiment based on the features produced by the DenseNet121 architecture. Both qualitatively and quantitatively evaluations demonstrate that our algorithm outperforms existing style transfer algorithms in terms of the sentiment transfer effects and input detail preservation.

VI. Acknowledges

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

REFERENCES

[1] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *ECCV*, 2018. 1, 2

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017. 1, 2

[3] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *ECCV*, 2018. 1, 2

[4] H. Tang, D. Xu, G. Liu, W. Wang, N. Sebe, and Y. Yan, "Cycle in cycle generative adversarial networks for keypoint-guided image generation," in *ACM MM*, 2019. 1

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *TIP*, 2004. 1, 2

[6] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015. 2, 3, 6

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012. 2, 4

[8] X. Alameda-Pineda, E. Ricci, Y. Yan, and N. Sebe, "Recognizing emotions from abstract paintings using non-linear matrix completion," in *CVPR*, 2016. 2

[9] A. Sartori, D. Culibrk, Y. Yan, and N. Sebe, "Who's afraid of itten: Using the art theory of color combination to analyze emotions in abstract paintings," in *ACM MM*, 2015. 2

[10] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *ACM MM*, 2010. 2

[11] X. Lu, P. Suryanarayan, R. B. Adams Jr, J. Li, M. G. Newman, and J. Z. Wang, "On shape and the computability of emotions," in *ACM MM*, 2012. 2

[12] J. Yuan, S. Mcdonough, Q. You, and J. Luo, "Sentribute: image sentiment analysis from a mid-level perspective," in *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 2013. 2

[13] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in *ACM MM*, 2014. 2

[14] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *ACM MM*, 2013. 2

[15] J. Yang, D. She, Y.-K. Lai, and M.-H. Yang, "Retrieving and classifying affective images via deep metric learning," in *AAAI*, 2018. 2

[16] J. Yang, D. She, and M. Sun, "Joint image emotion classification and distribution learning via deep convolutional neural network." in *IJCAI*, 2017. 2

[17] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in *AAAI*, 2015. 2

[18] J. Yang, D. She, Y.-K. Lai, P. L. Rosin, and M.-H. Yang, "Weakly supervised coupled networks for visual sentiment analysis," in *CVPR*, 2018. 2

[19] K. Song, T. Yao, Q. Ling, and T. Mei, "Boosting image sentiment analysis with visual attention," *Neurocomputing*, 2018. 2

[20] S. Zhao, Z. Jia, H. Chen, L. Li, G. Ding, and K. Keutzer, "Pdanet: Polarity-consistent deep attention network for fine-grained visual emotion regression," in *ACM MM*, 2019. 2

[21] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," *Neurocomputing*, 2019. 2

[22] Q. You, H. Jin, and J. Luo, "Visual sentiment analysis by attending on local image regions," in *AAAI*, 2017. 2

[23] L. Karacan, Z. Akata, A. Erdem, and E. Erdem, "Learning to generate images of outdoor scenes from attributes and semantic layouts," *arXiv preprint arXiv:1612.00215*, 2016. 2

[24] P. Sangkloy, J. Lu, C. Fang, F. Yu, and J. Hays, "Scribbler: Controlling deep image synthesis with sketch and color," in *CVPR*, 2017. 2

[25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017. 2

[26] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang, "Visual attribute transfer through deep image analogy," *arXiv preprint arXiv:1705.01088*, 2017. 2

[27] X. Huang and S. J. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *ICCV*, 2017. 2, 6

[28] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *CVPR*, 2016. 2

[29] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Controlling perceptual factors in neural style transfer," in *CVPR*, 2017. 2

[30] D. Kotovenko, A. Sanakoyeu, S. Lang, and B. Ommer, "Content and style disentanglement for artistic style transfer," in *ICCV*, 2019. 2

[31] Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, and J. Kautz, "A closed-form solution to photorealistic image stylization," in *ECCV*, 2018. 2, 6

[32] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *CVPR*, 2017. 2, 4, 6

[33] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *ICCV*, 2019. 2, 3, 6

[34] S. Bae, S. Paris, and F. Durand, "Two-scale tone management for photographic look," in *ACM Transactions on Graphics*, 2006. 2

[35] J. An, H. Xiong, J. Huan, and J. Luo, "Ultrafast photorealistic style transfer via neural architecture search," in *AAAI*, 2020. 2, 3, 6

[36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *CVPR*, 2017. 2

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 2

[38] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018. 2

[39] K.-N. Lianos, J. L. Schonberger, M. Pollefeys, and T. Sattler, "Vso: Visual semantic odometry," in *ECCV*, 2018. 2

[40] M.-J. Chen and A. C. Bovik, "Fast structural similarity index algorithm," *Journal of Real-Time Image Processing*, 2011. 3

[41] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *NeurIPS*, 2015. 3, 4

[42] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *CVPR*, 2016. 4

[43] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman, "Preserving color in neural artistic style transfer," *arXiv preprint arXiv:1606.05897*, 2016. 4

[44] E. Risser, P. Wilmot, and C. Barnes, "Stable and controllable neural texture synthesis and style transfer using histogram losses," *arXiv preprint arXiv:1701.08893*, 2017. 4

[45] S. Li, X. Xu, L. Nie, and T.-S. Chua, "Laplacian-steered neural style transfer," in *ACM MM*, 2017. 4

[46] Y. Li, N. Wang, J. Liu, and X. Hou, "Demystifying neural style transfer," *arXiv preprint arXiv:1701.01036*, 2017. 4

[47] T. Chen, W. Xiong, H. Zheng, and J. Luo, "Image sentiment transfer," *arXiv preprint arXiv:2006.11337*, 2020. 5

[48] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 5

[49] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," in *NeurIPS*, 2017. 6

[50] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *NeurIPS*, 2017. 5

[51] S. Mo, M. Cho, and J. Shin, "Instagan: Instance-aware image-to-image translation," *arXiv preprint arXiv:1812.10889*, 2018. 5

[52] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *CVPR*, 2018. 5

[53] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017. 5

[54] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *CVPR*, 2019. 5

[55] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018. 5

[56] T. Q. Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," *arXiv preprint arXiv:1612.04337*, 2016. 6

[57] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001. 7

[58] D. Coeurjolly, "Color transform via sliced optimal transfer," https://github.com/dcoeurjo/OTColorTransfer. 7