HAL open science

Superpixel-based matching of high-resolution deep features for color transfer

Hernan Carrillo, Michaël Clément, Aurélie Bugeau

▶ To cite this version:

Hernan Carrillo, Michaël Clément, Aurélie Bugeau. Superpixel-based matching of high-resolution deep features for color transfer. ORASIS 2021, Centre National de la Recherche Scientifique [CNRS], Sep 2021, Saint Ferréol, France. hal-03339667

HAL Id: hal-03339667
https://hal.science/hal-03339667

Submitted on 9 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Superpixel-based matching of high-resolution deep features for color transfer

Hernan Carrillo Michaël Clément Aurélie Bugeau

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

<prenom.nom>@labri.fr

Résumé

Dans cet article, nous proposons une nouvelle méthode pour la mise en correspondance de descripteurs haute résolution issus de CNNs *en utilisant des mécanismes d'attention. Cette méthode s'appuie sur une stratégie de* pooling *basée sur les superpixels pour calculer efficacement les similarités non-locales entre des paires d'images. Pour illustrer l'intérêt de ces nouveaux blocs méthodologiques, nous les appliquons au problème du transfert de couleur entre une image cible et une image de référence. Alors que les méthodes précédentes pour cette application peuvent montrer des problèmes de cohérence spatiale et colorimétrique, notre approche s'appuie sur une correspondance non-locale robuste entre des caractéristiques bas niveau de haute résolution. Enfin, nous soulignons l'intérêt de cette approche en montrant des résultats prometteurs en comparaison avec les méthodes de l'état de l'art.*

Mots Clef

Superpixels, mécanisme d'attention, transfert de couleur, descripteurs de haute résolution, correspondances non-locales

Abstract

In this article, we propose a new method for matching high-resolution feature maps from CNNs using attention mechanisms. This method relies on a superpixel-based pooling strategy to efficiently compute non-local similarities between pairs of images. To illustrate the interest of these new methodological blocks, we apply them to the problem of color transfer between a target image and a reference image. While previous methods for this application can suffer from poor spatial and color coherence, our approach tackles these problems by leveraging on a robust non-local matching between high resolution low-level features. Finally, we highlight the interest in this approach by showing promising results in comparison with state-of-the-art methods.

Keywords

Superpixels, Attention mechanism, Color transfer, High-resolution features, Non-local matching

1 Introduction

Color transfer aims at changing color characteristics of a target image by copying the ones from a reference image. Ideally, the result must reach a visually pleasant image, avoiding possible artifacts or improper colors. It covers various applications in areas such as photo enhancement, films post-production and artistic design.

Transferring the right colors requires computing meaningful similarities between the reference and the target images. These similarities must preserve important textures and structures of the target image. Most works on color transfer have focused on choosing the characteristics on which to compute similarities. These characteristics can be hand-crafted or learned using deep-learning methods. The first one extracts image features by relying on manually predefined descriptors (*i.e.*, HOG [4], SIFT [15]), however there is no guarantee that the descriptors are well suited for the task. The second solves this issue by learning the features from image dataset by leveraging on a training procedure, nonetheless features dimensionality increases enforcing the usage on low-resolution images. Features similarities can be matched using global information of the images (*i.e.*, color histograms); or local information such as matching small regions on the images (*i.e.*, cluster segmentation, superpixels decomposition). Finally, other efforts have been made on color fusion in case of color from several reference pixels are chosen to decide the target pixel color. In detail, color fusion frameworks rely on a weighted sum of spatial and color distances between neighbors reference pixels and a target image pixel.

In this paper, we compute similarities from high-resolution pre-trained deep learning features as this retains rich low-level characteristics. Due to dimensionality issue, we exploit existing superpixels extractor in order to match these high-resolution features and perform color transfer. The contributions are: 1) we propose super-features which encode deep learning features using superpixels decomposition; 2) we propose a robust non-local similarity between super-features using an attention mechanism; and 3) we build upon [8] and include these similarities in a non-local color fusion framework achieving promising results.

2 Related work

2.1 Superpixels

Exploiting superpixels representation allows finding interesting region's characteristics in images, such as color and texture consistency [1]. Many advantages can be derived using this type of decomposition, for instance, dimensionality reduction by grouping pixels with similar characteristics [25]. Additionally, this compact representation helps to overcome high computational costs on computer vision tasks such as object segmentation [24] or object localization [7]. However, the irregular form of the representation makes its usage difficult in computer vision tasks, especially the ones using deep learning approaches. But some works have proposed some representation to cope with this issue. For instance, [12] uses a superpixel label map as an input image to a neural network to extract meaningful information for clothing parsing application. [11] presents the SuperCNN as a deep neural network approach for salient object detection. It uses superpixels to describe two 1-D sequences of colors in order to reduce the computational burden. Nonetheless, neither of the existing approaches effectively encodes deep learning features for each superpixel.

2.2 Color transfer

Color transfer techniques can be classified into three classes: classic global-based methods, classic local-based methods, and deep learning methods.

Global methods consider global color statistics without any spatial information. It was initially introduced in [21] which uses basics statistical tools (i.e., mean, standard deviation) to match target and reference color information. [17, 29] extend color matching on different color spaces to find an optimal color mapping between the images. [6, 5] propose a global illuminant matching based on optimal transport color transfer for enforcing artifacts-free results. More complex methods such as [16] rely on Gaussian Mixture Models to create compressed signatures that ensure a compact representation of color characteristics between images. Nevertheless, these methods fail to ensure spatial consistency on resulting colors.

Local methods relies on spatial color mappings (i.e., segmentation, clustering) to match local regions of the target image and the reference image. [14] uses superpixel level style-related and style-independent feature correspondences. [2] implement a texture-based framework for matching local correspondence. Alternatively, [23] uses a probabilistic segmentation in order to impose spatial and color smoothness among local regions. Still, the method does not provide control over the matched superpixels. [8] overcomes this limitation by proposing a constrained approximate nearest neighbor (ANN) patches and a color fusion framework on superpixels. However, in this type of local methods target and reference images requires to share strong similarities.

Deep learning methods brings to the matching semantic-related characteristics from the target image and reference image. Recently [13] propose a deep neural network architecture that leverages on color histogram analogy for color transfer. The later uses a target and a reference histograms as input to exploit global histogram information over a target input image. And [10] relies on semantically meaningful dense correspondence between images. Nonetheless, this type of methods relies on pure semantic features (low-level features), which requires images from similar scene or instances.

2.3 Attention as a non-local operator

Non-local operators were introduced in image processing in [3] with the so-called Non-local means framework, initially used to filter out image noise by computing a weighted mean of all pixels in an image. Non-local means allow remote pixels to contribute to the filtered response, achieving less loss of details. It was then extended to non-local features matching for super-resolution [9], or inpainting [28], proving to achieve robust global features similarities.

Non-local similarities in neural network architectures were introduced in [26], called transformers architecture. A transformer is an end-to-end neural network approach that includes (self-)attention layers, which compute non-local similarities between multi-level feature maps. This type of architecture succeeds as a state-of-the-art method due to the capacity and flexibility of these attention blocks. The recent work [27] has bridged the gap between the self-attention mechanism [26] and non-local means. They stated that the self-attention mechanism captures long-range dependencies between deep-learning features by considering all features into the calculation.

Recently, the authors of [30] presented similarity calculation between different feature maps (target and reference images) based on attention mechanism. The principal drawback of such mechanism is the non-local operation, which has to be done on features with low dimensions due to computational overhead. In addition, low-resolution features usually do not carry sufficient information for calculating a robust pairwise similarity. For instance, deep features mainly carry high-level semantic information related to a precise application (*i.e.*, classification) that can be less relevant for high-resolution similarity calculation or matching purposes.

3 Method

We now present our color transfer method. It consists of three blocks: 1) super-feature extraction, 2) super-features matching, and 3) color transfer framework. In this section, all these steps will be illustrated on the target and reference images shown in Figure 1.

Our objective is to transfer colors from a reference RGB image $I_R \in \mathbb{R}^{H\times W\times 3}$ to a target RGB image $I_T \in \mathbb{R}^{H\times W\times 3}$. Concretely, this will be done by passing

Target image Reference image

Figure 1: Example of input images.

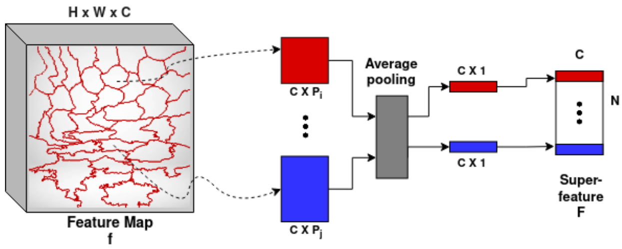


Figure 2: Diagram of our super-feature encoding proposal. This proposal is inputting a feature map of size $H \times W \times C$, in which each superpixel is extracted and encoded in vectors of size $C \times P_i$ pixels. Afterward, the vectors are pooled channel-wise and, finally, stacked in the super-feature matrix F with size $C \times N$ number of superpixels.

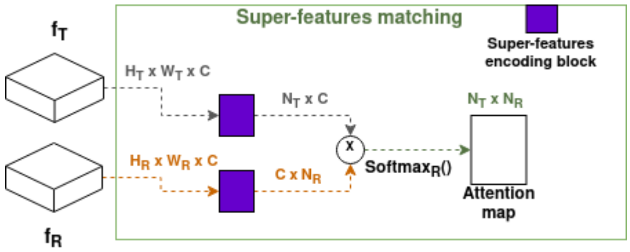


Figure 3: Diagram of our super-features similarity calculation. This layer takes a reference feature map f_R and a set of target feature map f_T as an input, and outputs an attention map at superpixels level by means of a non-local operation.

colors from I_R to I_T based on pairwise feature-related similarities.

3.1 Super-feature extraction

Let f_T and f_R be feature maps of the target image and reference image, respectively. In the following, we will consider features coming from pre-trained deep networks, but our methods can directly be applied to any other hand-crafted features. More precisely, we focus on features extracted at the first layers of a deep network, as they provide a long range of low-level features that suit diverse types of images. These feature maps then have high dimensions, typically the same size as the input image, times several channel's descriptors with $H \times W \times C$ where $C = 64$ or 128 for example.

A critical drawback of using high-resolution features for matching operations is the high computational complexity. Let the number of features in a feature map be $D = H \times W \times C$, then the complexity of the pixel-wise similarity computation is $\mathcal{O}\left(D^2\right)$. To solve this quadratic complexity problem, we implement an encoding layer based on superpixels representation. We first generate a superpixel map using a superpixels decomposition algorithm on the initial color images. Let us denote the target superpixel map by S_T, and the reference one by S_R. Each of these maps contains N_T and N_R superpixels respectively with P_i pixels each where i is the superpixel index. Next, we extract features of size $C \times P_i$ for each superpixel. These extracted features are then pooled by averaging channel-wise and stacked as a matrix of size $C \times N$ called super-feature F. Figure 2 illustrates this process. To sum up, the initial feature maps pass from size $H \times W \times C$ to super-feature encoding (F_T and F_R) of size $N_T \times C$ and, $N_R \times C$, making feasible operations such as correlation calculation between large deep neural networks features.

3.2 Super-features matching

Our super-features provide a compact encoding to compute the correlation between high-resolution deep learning features. Here, we take inspiration from the attention mechanism [30] to achieve a robust matching between target and reference super-features. The process is illustrated in Figure 3. Mainly, we exploit non-local similarities between the target and the reference super-features by computing the attention map as:

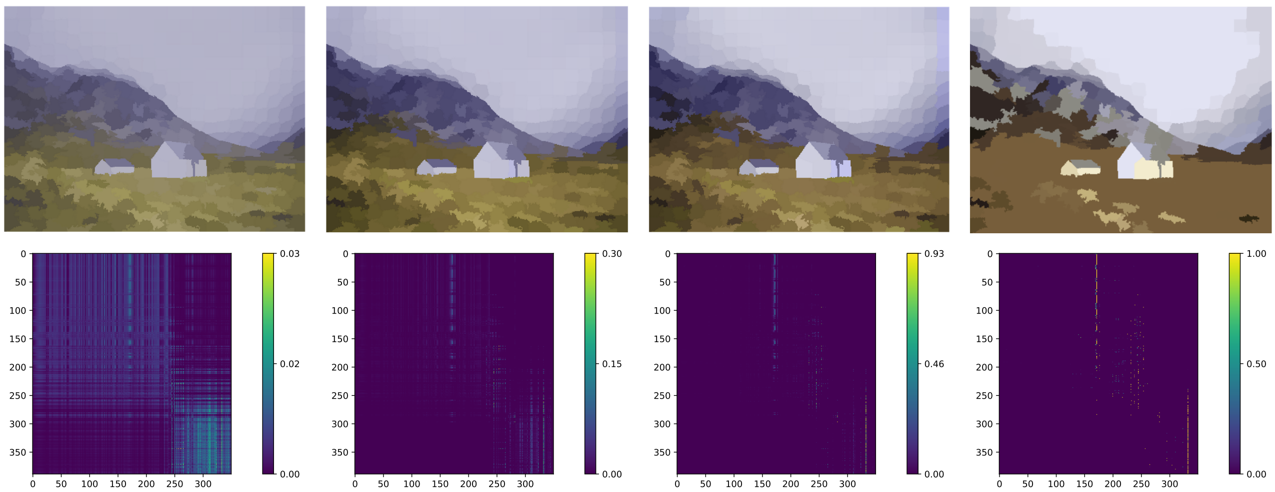
$$\mathcal{A} = \mathrm{softmax}_{\mathrm{R}}(\mathcal{M}_{TR}/\tau). \tag{1}$$

The $\mathrm{softmax}_{\mathrm{R}}$ operation normalizes row-wise the input into probability distributions, proportionally to the number of target superpixels N_R. Then, the matrix $\mathcal{M}_{TR}$ is a correlation matrix between the target and reference super-feature and is computed as:

$$\mathcal{M}_{TR}(i,j) = \frac{(F_T(i) - \mu_T) \cdot (F_R(j) - \mu_R)}{\|F_T(i) - \mu_T\|_2 \, \|F_R(j) - \mu_R\|_2} \tag{2}$$

where μ_T and μ_R are the mean of each super-feature. We found that this normalization keeps correlation values less sensitive to changes on τ for different images. The attention map (1) is the same non-local operator as the one proposed by Zhang *et al.* [30]. However, their computation requires low-level features due to the inherent complexity problem (as mentioned in Section 3.1).

We solve this complexity problem thanks to our super-feature encoding approach. Let $n = H \times W$ be the number of pixels in an image. Then the number of features in a deep learning feature map is $D = n \times C$ which translate in computational complexity as $\mathcal{O}(D^2) = \mathcal{O}(n^2 C^2)$. In contrast with our novel super-feature encoding, if we set the number of superpixels in the order of $\sqrt{n}$, then instead



$\tau = 0.1$ $\tau = 2 \times 10^{-2}$ $\tau = 9 \times 10^{-3}$ $\tau = 9 \times 10^{-4}$

Figure 4: Direct super-features matching using different τ values. Images on the first row depict our results for direct matching on superpixel-level. The second row represents ours attention maps for each of the images above. The attention map represents a probability distribution of the correlation between the target super-features (rows) and reference super-features (columns).

we rewrite with $D_s = \sqrt{n} \times C$, resulting in $\mathcal{O}(D_s^2) = \mathcal{O}(n \times C^2)$. As $C \ll n$ can be ignored, we go from quadratic to linear complexity operation. As a result, we can incorporate the correlation operation on large deep learning features from both target and reference images. Conversely, [30] can only rely on deep-level features, usually the bottleneck features (*i.e.*, $H/8 \times W/8 \times C$) for similarities calculation.

To match colors at superpixels level, we rely on the attention map $\mathcal{A}$ and the average of each superpixel color. Specifically, we apply our attention map as a soft-weight on the average colors, resulting then in a smooth colors correspondence (see Figure 4).

Figure 4 also shows the influence of temperature τ onto the superpixels attention map. We can see that the probability distribution is over-smoothed for larger values of τ (*i.e.*, $\tau = 0.1$), meaning that several reference super-features match one target super-feature. Otherwise, a small τ value means a hard one-to-one matching between a target and reference super-features (*i.e.*, $\tau = 9 \times 10^{-4}$).

3.3 Color fusion framework

Direct superpixel matching by averaging colors is not sufficient to obtain visually satisfying results. Image details are indeed lost at superpixel-level (*i.e.*, door, windows, etc., in Figure 4). Therefore we need to transfer color at pixel level from our superpixels matching.

For clarity in further equations, we denote the position and color centroids of a superpixel j in an image I as:

$$\bar{X}(j) = \frac{\sum_{p \in S(j)} p}{P_j}$$

and

$$\bar{I}(j) = \frac{\sum_{p \in S(j)} I(p)}{P_j}$$

respectively, where P_j is the number of pixels in superpixel j.

Inspired by the formulation of [8], we compute the new value $\hat{I}_t(p)$ of each pixel p of the target as a weighted average of reference superpixels representative colors:

$$\hat{I}_T(p) = \frac{\sum_{j=1}^{N_R} W(p,j)\bar{I}_R(j)}{\sum_{j=1}^{N_R} W(p,j)} \tag{3}$$

The weight matrix W depends firstly on the distance between pixel p and all target superpixels as in [8], and secondly, on our attention map:

$$W(p,j) = \sum_{i=1}^{N_T} d(p,i)\mathcal{A}(i,j). \tag{4}$$

The intuition behind the attention map is the addition of more relevant information about reference super-features into the transfer process. The distance between pixel p and superpixels centroids is computed over both positions and colors:

$$d(p,i) = e^{\left(-\frac{(V_T(p)-\bar{V}_T(i))^T \Sigma_i^{-1} (V_T(p)-\bar{V}_T(i))}{\sigma_g}\right)} \tag{5}$$

with position and color vectors being $V(p) = [p, I(p)]$ and $\bar{V}_T(j) = \left[\bar{X}_T(j), \bar{I}_T(j)\right]$, and the spatial and colorimetric covariances of pixels in superpixel i:

$$\Sigma_i = \begin{pmatrix} \delta_s^2 \operatorname{Cov}(p) & 0 \\ 0 & \delta_c^2 \operatorname{Cov}(I(p)) \end{pmatrix}. \tag{6}$$

Parameters δ_s and δ_s weight the influence of color and spatial information, respectively.

Finally, as in [8], after color fusion we apply a post-processing step using a color regrain algorithm [18], which

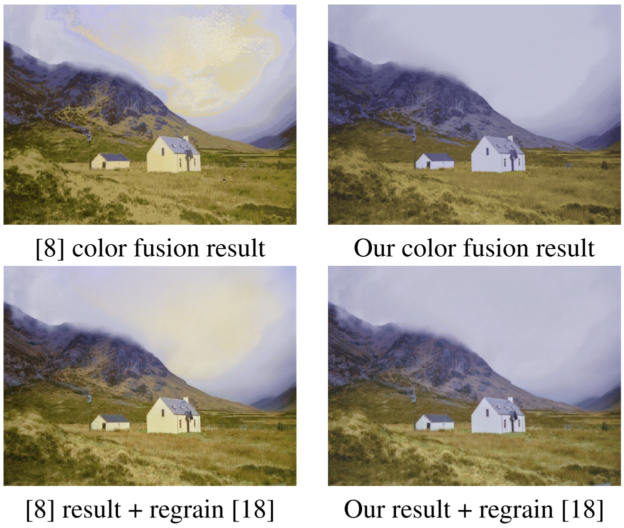


Figure 5: Color fusion framework results. A visual comparison between [8] and our result.

eventually matches the color distribution of I_R and the gradient of I_T. Figure 5 presents an example of our color transfer framework compared to the result of [8]. Visually, our results present better spatial consistency of colors. For instance, the sky on our results has a more natural smooth transition of colors. On the other hand, [8] results present a non-natural transition of colors (*i.e.*, yellow to blue).

4 Results

In this section, we first present the implementation details used to validate our method and then provide a detailed qualitative comparison between our results and four state-of-the-art approaches.

4.1 Implementation details

Superpixels representation is done using the SLIC algorithm [1], in which the number of superpixels depends on the actual size of the image. Experimentally, we set the number of superpixels as $3 \times \sqrt{n}$ where n is the number of pixels in the current image.

To build feature maps, we rely on a pre-trained VGG-19 [22] as our texture and color characteristics extractor, due to its simplicity and its 95.24% accuracy on the ImageNet Top-5 classes. However, our approach can work with other types of CNN architectures regardless its features dimensions.

In order to choose an optimal temperature τ value, we experimented on different images at distinct temperatures. Empirically, we obtain promising results using $\tau = 0.02$. In addition, all experiments have been run with $\delta_s = 10$ and $\delta_c = 0.1$, as [8] recommend to favor spatial consistency.

4.2 Comparison

We compare our method against four approaches: [19] which proposes an automated color transfer based on color distributions; [13] which implements a color transfer approach based on color histogram analogy using deep neural network; [20] that combines a gamut-based strategy with scene illumination; and [8] which implements the color fusion framework by leveraging on its proper superpixels decomposition. All four mentioned approaches had been considered state-of-the-art in color transfer, and have open-source codes for a fair comparison. Each method has been run with its default parameters.

Results comparing the four methods are shown in Figure 6. Our results (last column) have more visually pleasant colors and consistency in image texture, providing more realistic color transfer results with respect to the other methods.

For the first image (first row), [19, 20, 13] show over-saturation on the illumination of their results. Although this problem does not appear in [8] its result has a visible halo effect on the colors of the sky (darker blue to light blue). On the second image, most of the results fail to transfer the blue in the sky of the desert to the sea onto the target image, except for [8], which does it partially at the cost of a saturate orange on the reflection of the water. Our result for the second image corrects this issue making a more pleasant result. Lastly, our third color transfer result shows that fine details such as the line that divides the road is caught, and distinguishes from other textures. Finally, all compared methods achieve at least one over-saturated result. However, our method ensures more realistic but opaque colors. This happens for two reasons: 1) we rely on average colors from superpixels regions that lead to less saturated colors results; and 2) the attention temperature τ, which smoothes the weights. We believe our results are more realistic, and we will study in future works if saturated results keep this realism.

5 Conclusions

This paper proposed a new method for color transfer by leveraging deep learning features and superpixels. At the core of the method is our novel super-features matching that uses high-level deep learning features from both target and reference images. On top of that, we update the color fusion framework proposed by [8] to consider our attention map, which provides texture and color knowledge from the reference image onto the final color transfer step. Finally, our method achieves more visually consistent and realistic results in comparison to the four state-of-the-art methods considered. Work is underway on adapting the method for transferring color from a reference color image to a grayscale target image. Another essential extension will be a generic method to handle and combine both low and high dimensional features.

6 Acknowledgements

This study has been carried out with financial support from the French Research Agency through the PostProdLEAP project (ANR-19-CE23-0027-01).

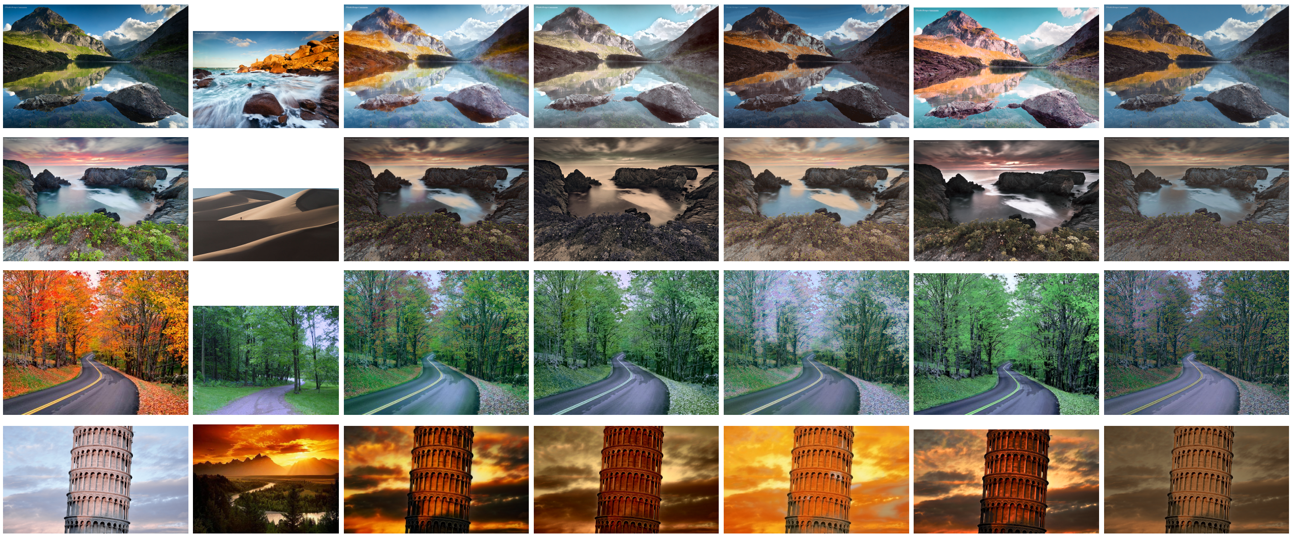


Figure 6: Comparison of color transfer results. We compare our method with four different state-of-the-art approaches: [19] color distribution grading, [20] gamut-based color transfer, [8] color fusion based on superpixel representation and, [13] deep learning-based color histogram analogy. Finally, our method produces more realistic and pleasant results.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[2] B. Arbelot, R. Vergne, T. Hurtut, and J. Thollot. Local texture-based color transfer and colorization. *Computers & Graphics*, 62:15–27, 2017.

[3] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Conference on Computer Vision and Pattern Recognition*, 2005.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Conference on Computer Vision and Pattern Recognition*, 2005.

[5] S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, and J.-F. Aujol. Regularized discrete optimal transport. In *Scale Space and Variational Methods in Computer Vision*, 2013.

[6] O. Frigo, N. Sabater, V. Demoulin, and P. Hellier. Optimal transportation for example-guided color transfer. In *Asian Conference on Computer Vision*, 2015.

[7] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. In *International Conference on Computer Vision*, 2009.

[8] R. Giraud, V.-T. Ta, and N. Papadakis. Superpixel-based color transfer. In *International Conference on Image Processing*, 2017.

[9] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *International Conference on Computer Vision*, 2009.

[10] M. He, J. Liao, D. Chen, L. Yuan, and P. V. Sander. Progressive color transfer with dense semantic correspondences. *ACM Transactions on Graphics*, 38(2), 2019.

[11] S. He, R. Lau, W. Liu, Z. Huang, and Q. Yang. Supercnn: A superpixelwise convolutional neural network for salient object detection. *International Journal of Computer Vision*, 115, 04 2015.

[12] A. Ihsan, L. Chu Kiong, S. Naji, and M. Seera. Superpixels features extractor network (SP-FEN) for clothing parsing enhancement. *Neural Processing Letters*, 51:2245–2263, 2020.

[13] J. Lee, H. Son, G. Lee, J. Lee, S. Cho, and S. Lee. Deep color transfer using histogram analogy. *The Visual Computer*, 36(10):2129–2143, 2020.

[14] J. Liu, W. Yang, X. Sun, and W. Zeng. Photo stylistic brush: Robust style transfer via superpixel-based bipartite graph. In *International Conference on Multimedia and Exposition*, 2016.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[16] N. Murray, S. Skaff, L. Marchesotti, and F. Perronnin. Toward automatic and flexible concept transfer. *Computers & Graphics*, 36(6):622–634, 2012.

[17] F. Pitié and A. Kokaram. The linear monge-kantorovitch linear colour mapping for example-based colour transfer. In *European Conference on Visual Media Production*, 2007.

[18] F. Pitié, A. Kokaram, and R. Dahyot. Towards automated colour grading. In *IEEE European Conference on Visual Media Production*, 2005.

[19] F. Pitié, A. C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1–2):123–137, 2007.

[20] N. H. M. Rang, S. J. Kim, and M. S. Brown. Illuminant aware gamut-based color transfer. *Computer Graphics Forum*, 33, 2014.

[21] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *ACM Transactions on Graphics*, 21(5):34–41, 2001.

[22] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 2015.

[23] Y.-W. Tai, J. Jia, and C.-K. Tang. Local color transfer via probabilistic segmentation by expectation-maximization. In *Conference on Computer Vision and Pattern Recognition*, 2005.

[24] J. Tighe and S. Lazebnik. Superparsing: Scalable nonparametric image parsing with superpixels. In *European Conference on Computer Vision*, 2010.

[25] M. Van den Bergh, X. Boix, G. Roig, and L. Van Gool. SEEDS: Superpixels extracted via energy-driven sampling. *International Journal of Computer Vision*, 111(3):298–314, 2015.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.

[27] X. Wang, R. Girshick, A. Gupta, and K. He. Non-local Neural Networks. In *Conference on Computer Vision and Pattern Recognition*, 2018.

[28] Y. Wexler, E. Shechtman, and M. Irani. Space-time video completion. In *Conference on Computer Vision and Pattern Recognition*, 2004.

[29] X. Xiao and L. Ma. Color transfer in correlated color space. In *International Conference on Virtual Reality Continuum and its Applications*, 2006.

[30] B. Zhang, M. He, J. Liao, P. V. Sander, L. Yuan, A. Bermak, and D. Chen. Deep exemplar-based video colorization. In *Conference on Computer Vision and Pattern Recognition*, 2019.