Find me a sky : a data-driven method for color-consistent sky search & replacement

✉ Saumya Rawat*, Siddhartha Gairola*, Rajvi Shah, and P J Narayanan

Center for Visual Information Technology, KCIS, IIIT Hyderabad, India
saumya.rawat@students.iiit.ac.in

Abstract. Replacing overexposed or dull skies in outdoor photographs is a desirable photo manipulation. It is often necessary to color correct the foreground after replacement to make it consistent with the new sky. Methods have been proposed to automate the process of sky replacement and color correction. However, many times a color correction is unwanted by the artist or may produce unrealistic results. We propose a data-driven approach to sky-replacement that avoids color correction by finding a diverse set of skies that are consistent in color and natural illumination with the query image foreground. Our database consists of $\sim$ 1200 natural images spanning many outdoor categories. Given a query image, we retrieve the most consistent images from the database according to L_2 similarity in feature space and produce candidate composites. The candidates are re-ranked based on realism and diversity. We used pre-trained CNN features and a rich set of hand-crafted features that encode color statistics, structural layout, and natural illumination statistics, but observed color statistics to be the most effective for this task. We share our findings on feature selection and show qualitative results and a user-study based evaluation to show the effectiveness of the proposed method.

1 Introduction

With the ubiquity of smart phone cameras, photography has become a democratized hobby with millions of photos uploaded to social media platforms like Instagram, Flickr, Facebook every day. Along with this comes the need for sharing perfect photographs, however, the captured shots are often unattractive due to undesirable backgrounds, occlusions, poor lighting or exposure, motion blur, lack of smile, presence of eye blinks, etc. In recent years, many methods are proposed for a number of automatic photo enhancements. This paper focuses on the problem of automatic sky-replacement.

Sky is often the hardest part of the scene to perfect in outdoor photography. Depending upon the geographic location and weather conditions, sky could persistently be gray and dull, or too bright. Even when the sky is perfect blue with white clouds and looks beautiful to the naked eye, it most often gets washed out in a single exposure shot captured with a standard smart-phone camera. Professional outdoor photographers often prefer the golden hour (when sun is closer to the horizon) or use specifically designed filters and polarizers to overcome this problem. Multi-exposure (HDR) photography can alleviate this problem to some extent, however, not much can be done if at the time of capture sky is just dull.

[*] Both authors have contributed equally to this work.
Project page : *https://cvit.iiit.ac.in/research/projects/cvit-projects/findmeasky*

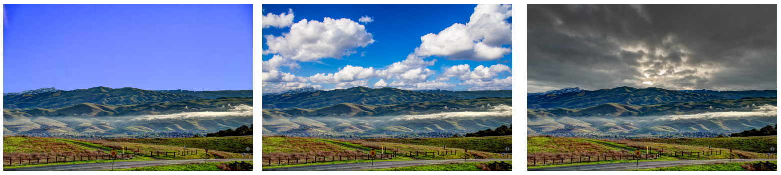

Fig. 1: For a query image with a dull sky (left), examples of consistent (middle) and inconsistent (right) sky replacements.

Professional digital artists, perfect the bad-sky photographs by manually replacing the original sky with a desirable one and performing a series of interactive corrections to make the sky and the foreground consistent with each other while keeping the final composite 'plausible'. This is a non-trivial and time consuming edit that is too cumbersome for a naïve user to perform. Recently, [10] proposed an automatic method for sky-replacement that performs semantic-aware color transform on the foreground to achieve natural looking composites. However, color-correction is not always desirable. Hence, we propose a different approach to sky-replacement that avoids or minimizes the need for post-replacement color corrections.[6]

Our approach is data-driven and centered around the idea of 'compatible' sky-search. Given a query image with a problematic sky, our method first finds images with similar foregrounds and natural illumination. It then creates candidate composites by replacing the query image sky with the retrieved image skies and ranks the composites based on realism and diversity. The user is finally presented with the top-k candidate composites as replacement outcomes without color transfer thereby retaining the natural color composition of the foreground in the original image. We demonstrate the effectiveness of our method with qualitative results and a comprehensive user study. Figure 2 summarizes the proposed system with a block diagram.

For retrieving compatible yet useful images, we curated a dataset of 1246 outdoor images spanning many outdoor categories with interesting skies from ADE20K dataset [19] and the dataset of [11]. To achieve compatible sky-search, we use an ensemble of hand-crafted features such as Color Statistics (Correlated Color Temperature (CCT), Luminance, and Saturation histograms), GIST [5], Bag of Words[9], and natural illumination statistics [4] (represented as a probability map of sun position in the sky); as well as CNN features (pre-trained). These features encode rich information about color distribution, structural layout, semantics, and natural illumination. We finally select the color statistical features, as we found based on an ablation study that the composites produced using the retrieval results with these features were most realistic. We evaluate the composite images using RealismCNN [20] – a discriminative model trained to predict realism of an image. Section 3 explains the data collection, feature selection, and re-ranking based on realism and diversity in detail.

To summarize, our contributions are the following, (i) We present a novel pipeline for compatible-search based sky-replacement that is a useful alternative or prelude to automatic color transfer based methods. (ii) We curated a large database of outdoor images with interesting skies and evaluated usefulness of a large number of features for this task. Our findings along with the database would be useful to the community for future research in this direction.

2 Related Work

Automatic sky-search and sky-replacement: Tao et al. [10] proposed an interactive search system using a set of semantic sky attributes (category, layout, richness, horizon, etc.) and showed how it can be used for controllable sky replacement. However the sky segmentation and consequently horizon detection introduce errors in sky replacement. Tsai et al. [11] proposed a data-driven sky search scheme based on semantic layout of the input image. To re-compose the stylized sky with the original foreground naturally, an appearance transfer method is developed to match statistics locally and semantically. However, the color transfer algorithm is linked with label matching between the source and the target which adds both complexity and a limitation on the kind of source images that can be used. Also, color transfer may be undesirable and may introduce artefacts in the foreground regions. In contrast, we do not rely on similar sky replacement methods and also do not need to use appearance transfer methods.

Realistic image composition: Much work has been done for realistic image composition [15] and for evaluating realism of composites[14, 16]. Lalonde and Efros [3] propose an object insertion technique that searches for objects that are consistent with the input photograph in terms of camera orientation, lighting, resolution, etc. and uses feature based assessment of composite realism. Xue et al. [17] determine the key statistical measures that influence the realism of a composite and then adjust these in a given query composite automatically using a data-driven algorithm. In this work, we leverage the implicit correlation between background and foreground regions in natural images for compatible sky-search that lead to more realistic composites.

3 Proposed system

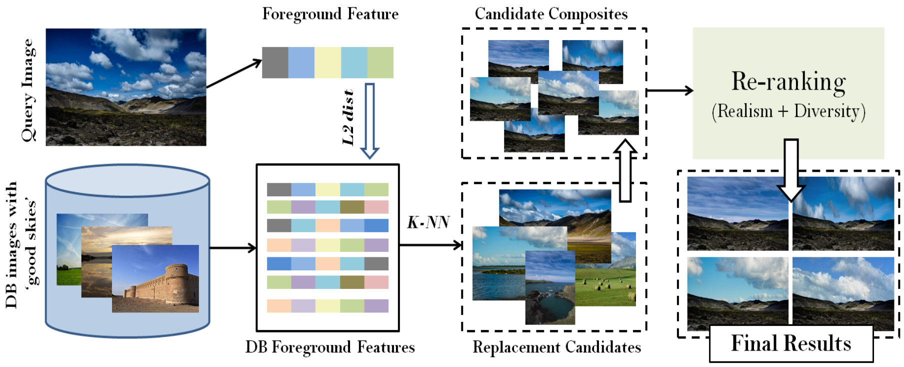

Fig. 2: An overview of our sky-replacement pipeline. Candidate composites are created using skies from database images with most similar foregrounds. Final composites are re-ranked to maximize realism and diversity of the presented set.

The motivation behind our sky replacement method is to find naturally consistent yet interesting skies for a query image. Our system is based on the following hypothesis. Given two images, (i) if their foreground regions are similar (in color, layout, and semantic makeup), and (ii) if the estimated natural illumination (predicted positions of the sun in the sky) is similar, swapping their skies would lead to highly realistic composites that wouldn't need foreground color correction. This hypothesis is validated with experiments (discussed later). We first curate a database of outdoor images with interesting and aesthetically appealing skies along with their foreground masks. We represent each database image with image features corresponding to its foreground region and illumination. Similarly, given a query image (and its foreground mask), we compute its foreground features and natural illumination. For each query, we retrieve the top-K nearest neighbor images from the database based on the L_2 distance in feature space and use the sky regions in these images as viable candidates for replacement. We evaluate all candidate composites for realism and diversity and re-rank the candidates to provide most realistic yet diverse alternatives to the query image. This procedure is outlined in Figure 2.

3.1 Database collection

The database of 1246 images used with the proposed system consists of 415 Flickr images with diverse skies (collected by [11]) and 831 outdoor images curated from the ADE20K Dataset[19]. ADE20K dataset consists of $\sim$ 22K images with 150 semantic categories like sky, road, grass. The images with sky category were first filtered to a set of $\sim$ 6K useful images for which the sky region made $>40\%$ of the total image. These images were manually rated between 1 to 5 for interestingness and aesthetic appeal of the skies by two human raters and only the images with average scores higher than 3 were added to the final database.

3.2 Feature representation

To find images with a similar foreground make-up and illumination, we performed experiments with a rich set of hand-crafted and pre-trained CNN features and found the feature based on color statistics to be equal or more effective than pre-trained CNN features for this task. In this section, we briefly introduce the features used for (i) foreground representation, and (ii) natural illumination representation, and explain the ablation on effectiveness of individual features and their combinations in the next section.

Foreground Features

Color statistics: Xue et al. [17] studied the relation between the background and foreground regions for realistic composition using various 2D statistical measures and identified correlated color temperature (CCT), luminance, and saturation to be the most significant measures in determining realism of a composite. We use this finding and represent the image foreground using histograms of these color statistics computed at every pixel (using [13]).

Bag of Visual words and GIST: Hand-crafted features such as Bag of visual words (BoW) [9] and GIST have been popularly used for measuring object-level and scene-level similarities between images. For our task, BoVW features are computed by quantizing densely extracted local descriptors (like SIFT) from foreground region of an image into a large visual vocabulary and building a normalized histogram of these word occurrences. GIST features are designed to capture spatial envelope of the scene and use histogram representation of gabor filter responses applied at multiple scales and orientations. We use VLFeat library [12] to extract BoW and GIST features.

Pre-trained CNN features: Image descriptors computed using convolutional neural networks (CNNs) pre-trained on large data such as ImageNet have proven to be very effective for a number of visual understanding tasks. The success of these features can be attributed to implicit learning of spatial layout and object semantics at later layers of the network from very large datasets. We use two different pre-trained networks, (i) VGG19 architecture [8] trained on ILSVRC-2012 (ImageNet) dataset, and (ii) VGG16 architecture trained on Places205 dataset [18], and extract two variants of CNN features. With both architectures, we use the output of FC7 (fully-connected) layer (4096 dim.) as feature representation. Between these two, ImageNet pre-trained CNN features performs better. We did not fine-tune these networks for our task due to lack of labeled data.

Illumination Features

Sun position & visibility: Apart from foreground similarity, images with illumination similar to the query would be better candidates for sky replacement. We compare the sun position in the sky estimated using [4]. This method estimates a probability distribution over sun position in the sky (azimuth and zenith angles) and visibility using a combination of weak cues (sky pixel intensities, cast shadows on ground, vertical surface shading) and a data-driven prior.

3.3 Candidate search and composition

Candidate Search: The query and the candidates are compared using a combination of foreground distance (d_{fg}) and the sun position distance (d_{il}) as follows,

$$d(I_q, I_c) = d_{fg}(I_q, I_c) + \alpha \, d_{il}(I_q, I_c) \tag{1}$$

Foreground features are compared using L_2 distance. For comparing illumination, instead of comparing two probability distributions, we directly compute the angular distance (zenith and azimuth) between the query and the candidate images. If the highest probability is below 0.5, the parameter α is 0, we do not consider the illumination distance as reliable and discard it otherwise α is 1. Distances are normalized between 0 and 1.

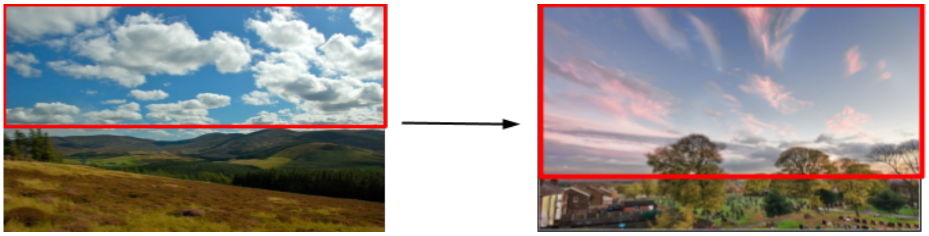

Fig. 3: An overview of the sky replacement step.

Composition: The database images are stored with an alpha mask corresponding to the sky/foreground segmentation. We assume the availability of alpha mask for query image also. [11] explain an automatic method to obtain accurate sky segmentation. Alternatively a semi-automatic method [7] can be used to obtain a reliable alpha mask for the query image. Given the query and the candidate images with corresponding segmentation masks, we first crop the tightest rectangle consisting only of the sky pixels from the candidate image and scale it to match the size of the maximum bounding rectangle of the query image. We then replace the query image sky patch by the scaled candidate sky patch as illustrated in Figure 3 and perform laplacian pyramid based blending [1] along the seam to reduce composition artifacts.

3.4 Feature Selection based on composite realism

Given a query, the ideal feature is the one that yields candidate images with most suitable skies for replacement. Suitability of an image for this task is determined by perceived realism and aesthetic appeal of the final composite. These properties are highly subjective and hence obtaining ground-truth rankings/ratings for a large number of query images requires extensive human annotation effort. Recently, [20] trained a discriminative model to predict realism of an image (RealismCNN). While, this is not an accurate indicator of ‘goodness’ of a candidate for our task, it is a useful alternative to validate usefulness of the features in absence of any ground-truth/baseline. We created a validation set of 100 query images for this ablation study. For each query, we retrieved the top-100 candidates using L_2 distance of the five foreground features and also using a combination of foreground and sun position distances. This leads to 100 composites per query per feature (100K composites per feature). Using the predictive model of [20], we obtain a realism score for each composite.

Figure 4 shows the running average of realism scores for incremental subsets of top-K composites (10%, 20%, ..., 100%). As discussed before, our hypothesis is that using skies of images with most similar foregrounds and/or illumination would lead to most realistic composites. If this hypothesis is valid, with increase in value of K, average realism score of the top-K composites should be decreasing. This trend can be observed for all features, validating our hypothesis. Among all foreground features, color statistics feature yields the highest average realism scores, CNN feature is a close second (ImageNet pre-trained). We study the effect of these two features combined with illumination feature (sun positions) as per equation 1. While the average scores

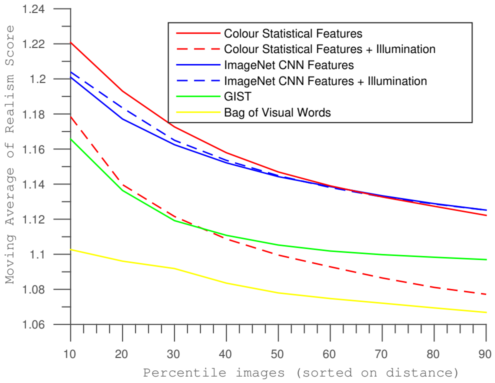

Fig. 4: Ablation study : Running average of realism scores for composites sorted on feature distances

drop for combined illumination and color features, these features are helpful to avoid physically implausible composites. But since the performance is significantly lower we finally only use color statistic features for finding the suitable candidate skies.

3.5 Re-ranking for realism and diversity

While the candidates obtained using feature based distances are compatible and the resulting composites are realistic, presenting all composites to the user is unnecessary and often undesirable. Many composites can potentially be redundant if the replaced sky is similar to the query and/or to other composites. We propose to select a small and diverse subset of highly realistic composites. To achieve this, the composites are re-ranked based on the realism score (RealismCNN) and a diversity measure. This is done by casting this problem to a max-sum diversification objective and optimizing this objective using a facility dispersion algorithm as proposed by [2].

For relevant and diverse retrieval, we wish to select a subset that maximizes total relevance ($\sum w$) and total dissimilarity ($\sum d$). Consider U is the set of all candidate composites for a query image I_q and $S \subseteq U$ is the desired subset. The bi-criteria objective ($f(S)$) that achieves this can be given by Equation 2 [2] (where $\lambda > 0$ is a trade-off parameter).

$$f(S) = (k-1)\sum_{u\in S} w(u) + 2\lambda \sum_{u,v\in S} d(u,v) \tag{2}$$

$$d'(u,v) = w(u) + w(v) + 2\lambda d(u,v) \tag{3}$$

To recast the objective as *max-sum dispersion* (that maximizes sum of all pairwise distances in the subset S), [2] introduces a new pairwise distance given in Equation 3. For our task, we want the composite to be realistic and the sky regions to have comparable aspect ratios hence, (i) relevance w for each composite is a product of it's min-max normalized realism score and the scale factor (i.e scaling applied to the candidate sky patch), and (ii) the dissimilarity d is the L_2 distance between between two *sky regions* in color feature space.

4 Results and Discussion

Our system is implemented in MATLAB with binary bindings for realism evaluation and blending. Currently, the code is not optimized for performance and takes around a minute to produce 100 candidates for a query image, of which, we show the top-4. To evaluate the effectiveness of our method, we show qualitative results for a few query images and discuss findings of the user-study based evaluation conducted for a larger query set.

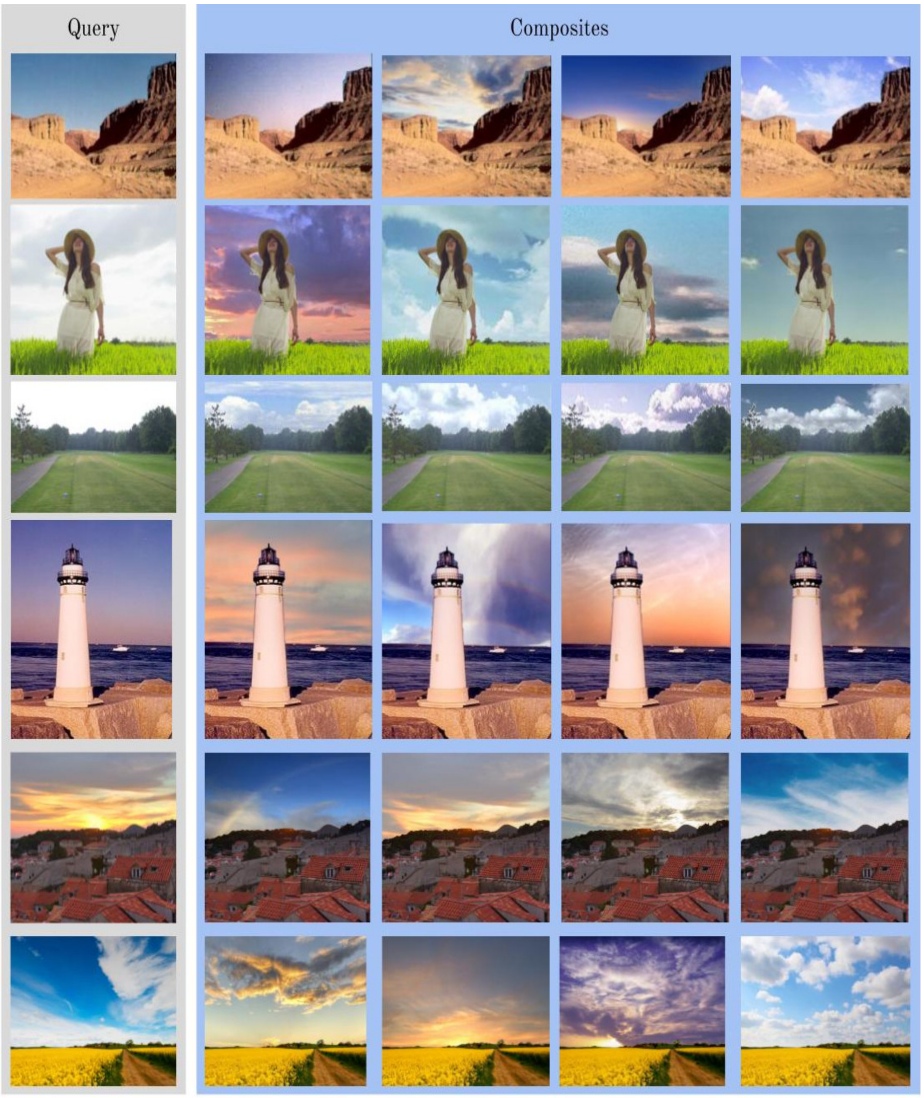

Fig. 5: Example results of our diverse and compatible sky-replacement system

Qualitative results: Figure 5 illustrates the 4 best composites for the query images on the top. The query images shown include a variety of scene types and configurations such as aerial/ground shots, presence/absence of foreground objects (person, tower), dull/interesting skies. It can be seen that for all queries, the composites are diverse, natural looking, and aesthetically appealing. Figure 6 shows the usefulness of the re-ranking algorithm. The images before re-ranking have similar backgrounds to the input image. But after re-ranking we get images which are both diverse and relevant. Figure 8 compares the results from the given pipline and the results given by [11]. The comparison clearly shows that our method produce results which are similar in aesthetic appeal. Figure 7 depicts the failure of the color transfer techniques used by [11] as the specularity and reflection from the roofs in the houses is clearly visible. There is no need for such correction in our method as it chooses skies that are already compatible with the foreground of the input image.

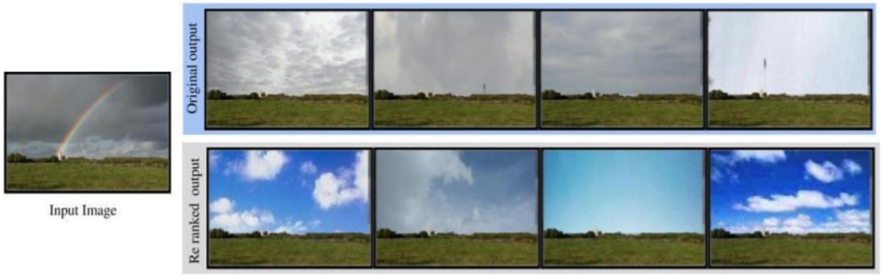

Fig. 6: Example illustrating the efficacy of the re-ranking.

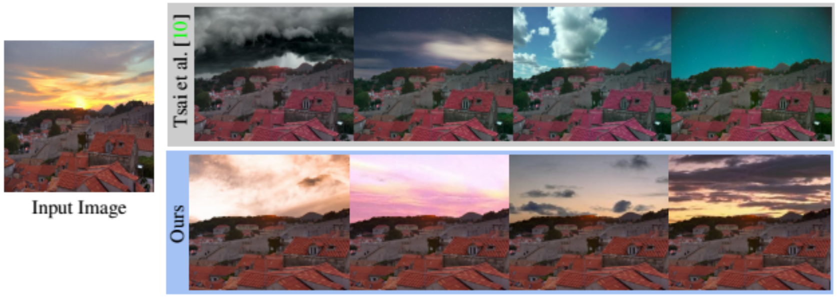

Fig. 7: Failure of colour transfer methods in the in-house implementation of [11] as compared with our method which chooses skies that are already compatible with the foreground.

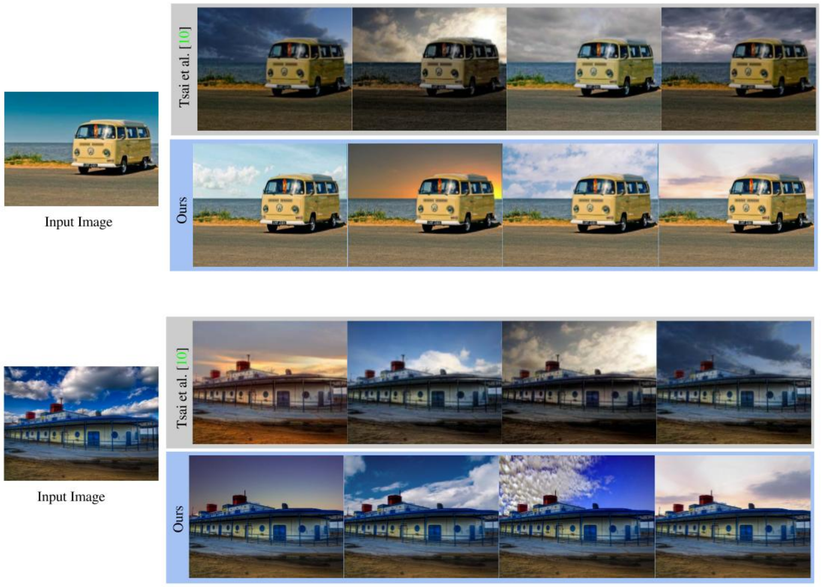

Fig. 8: Comparison with existing state of the art method, [11]. We tested our model on the same input image, the results obtained are just as aesthetically appealing using a completely different pipeline.

	min	max	mean	median
$\mathcal{R}_q >$ all $\mathcal{R}_c$	0%	52%	12.72%	8.6%
any $\mathcal{R}_c \geq \mathcal{R}_q$	48.48%	100%	87.32%	91.4%
any $\mathcal{R}_c > \mathcal{R}_q$	10.5%	81.67%	43.38%	43.31%

Table 1: Statistics on user preferences

User-study evaluation: To assess the performance of our replacement system, we conducted a user study where we asked the users to rate groups of images based on their naturalness and aesthetic appeal. Each group included a query image and top-3 composites in a randomized (and anonymous) fashion. The user study was conducted for a set of 30 query groups and each group was rated by at least 40 participants. The participants belonged to age group 20 to 35 and had varying degrees of photography and composition expertise, with a larger segment self-identifying as amateur or casual photographers. Each image was rated between scores 1 to 5 which correspond to 'very bad', 'bad', 'okay', 'good', and 'very good' descriptions. In absolute terms, the median score (across users and queries) for the original image is 2.82 (below 'okay') while for the composites, it is 3.12 (above 'okay') indicating that the composites were perceived to be equally or more attractive than the original images, Relatively, 83.33%

of the times at least one out of three composites received a rating strictly higher than the query image indicating preferable aesthetic appeal of the composites. We also report statistics on the fraction of times a query image I_q is rated $>$, $=$, $<$ any of the composite images I_c in Table 1. It shows user agreement for various cases, e.g. for the criteria any $\mathcal{R}_c > \mathcal{R}_q$ (where $\mathcal{R}_i$ is rating of an image i), the worst performing query set (column corresponding to min) 10.5% users agree, the best set has 81.67% users in agreement, and on average over all query sets 43.38% users agree. The survey results clearly indicate merit in our replacement system.

Failure Cases: Figure 9 shows a few cases where our pipeline fails. Figure 9 (i) illustrates that like any composition system, success of our system also assumes accurate segmentation and incorrect segmentation can lead to inconsistent composites. In case of scenes with specular surfaces like in (ii–iii), inconsistent reflection of the sky can lead to unnatural looking compositions.

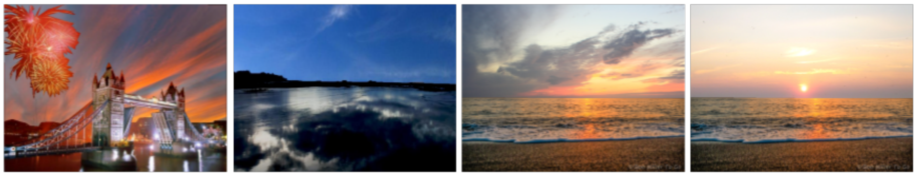

Fig. 9: Failure cases, from left, (i) segmentation error, (ii) inconsistent sky reflection in water, (iii) bright (sun) spot, (iv) better composition achieved with use of illumination map.

5 Conclusions

In this paper, we proposed a data-driven method that given a query image produces interesting and realistic composites with different skies without using color transfer as a post-processing step. To achieve interesting replacements, we curated a new dataset of outdoor images with interesting skies. To achieve realism without color transfer, we proposed a foreground similarity hypothesis and validated it using a realism prediction model. We also experimented with a variety of image based features for this task and observed color statistical features to be very effective. We further showed a re-ranking technique to achieve both realism and diversity in the final subset presented to the user. The effectiveness of our method is evaluated by conducting a thorough user study. In future, we would like to explore an unsupervised learning based alternative to our selection pipeline and also explore generative formulation of sky-replacement problem.

Bibliography

[1] Peter J. Burt and Edward H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2(4), October 1983.

[2] Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. In *Proc. ACM WWW*, 2009.

[3] J.-F. Lalonde and A.A. Efros. Using color compatibility for assessing image realism. In *Proc. IEEE ICCV*, 2007.

[4] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Estimating natural illumination from a single outdoor image. In *Proc. IEEE ICCV*, 2009.

[5] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42(3), 2001.

[6] Saumya Rawat, Siddhartha Gairola, Rajvi Shah, and P. J. Narayanan. Find me a sky : a data-driven method for color-consistent sky search & replacement. In *The 24th International Conference on Multimedia Modeling (MMM 2018), Bangkok, Thailand*, 2018.

[7] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. ”grabcut”: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3), 2004.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[9] Josef Sivic and Andrew Zisserman. Video google: A text retrieval approach to object matching in videos. In *Proc. IEEE ICCV*, 2003.

[10] Litian Tao, Lu Yuan, and Jian Sun. Skyfinder: Attribute-based sky image search. *ACM Trans. Graph.*, 28(3), 2009.

[11] Y.-H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, and M.-H. Yang. Sky is not the limit: Semantic-aware sky replacement. *ACM Trans. Graph.*, 35(4), 2016.

[12] A. Vedaldi and B. Fulkerson. VLFeat - an open and portable library of computer vision algorithms. In *ACM MM*, 2010.

[13] J. Wågberg. *OptProp: Matlab Toolbox for Calculation of Color Related Optical Properties : Version 2.1*. FSCN-rapport. 2007.

[14] Dong Wang, Weijia Jia, Guiqing Li, and Yunhui Xiong. Natural image composition with inhomogeneous boundaries. In Yo-Sung Ho, editor, *Advances in Image and Video Technology: Pacific Rim Symposium (PSIVT)*, 2012.

[15] Jue Wang and Michael F. Cohen. Image and video matting: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(2), 2007.

[16] Bing-Yi Wong, Kuang-Tsu Shih, Chia-Kai Liang, and Homer H. Chen. Single image realism assessment and recoloring by color compatibility. *IEEE Trans. Multimedia*, 14, 2012.

[17] Su Xue, Aseem Agarwala, Julie Dorsey, and Holly Rushmeier. Understanding and improving the realism of image composites. *ACM Trans. Graph.*, 31(4), 2012.

[18] Bolei Zhou, Aditya Khosla, Àgata Lapedriza, Antonio Torralba, and Aude Oliva. Places: An image database for deep scene understanding. *CoRR*, abs/1610.02055, 2016.

[19] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proc. IEEE CVPR*, 2017.

[20] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A. Efros. Learning a discriminative model for the perception of realism in composite images. In *Proc. IEEE ICCV*, 2015.