Deep Image Harmonization

1 Data Acquisition

To train an end-to-end deep CNN, image pairs containing composite images and corresponding harmonized images are required as the input and output ground truth. But unfortunately, it's not so easy to get such image pairs for image harmonization task. Generating high-quality harmonized images requires extensive efforts of skilled expertise, for which reason it's not feasible to generate large-scale training data.

To tackle this problem, [1] provides an inspiring way to generate such image pairs. They start from a real image treated as the output ground truth of the network. Then select a region, apply color transfer to edit its appearances, generating an edited image working as the input composite image. This approach to making image pairs guarantees the ground truth images are always realistic, freeing us from time-consuming expertise editing.

But it also has some drawbacks in the way of generation. To transfer color between images, [1] uses histogram matching. As we introduce in the following, it's a non-parametric method within decorrelated color space and belongs to one of the four branches in our color transfer taxonomy, which is not sufficient to cover a wide range of representative methods. Besides, it changes the luminance and color temperature only. Though applying different transfer parameters for these two statistics, it is not enough to simulate real cases.

We follow the idea of [1] and make some improvements. We start with a real image with a segmentation mask, treating the real image as the ground truth. Then we randomly select a relatively large region (e.g., an object or a scene) with the help of segmentation mask. When we come to the step to edit its appearance, instead of using histogram matching, we do a survey and select four color transfer method candidates to represent all methods in color mapping field. After applying one of them, the edited image will be used as the composite input of our network. And in the following, we introduce the details about color transfer methods taxonomy and the data acquisition process.

1.1 Classification of color transfer methods

Color transfer or color mapping is a class of approaches that aim to apply the color palette or style from the reference to the target image.

When generating composite images, the pair of the target image and reference image generally doesn't come from similar scenes or the same scene. As we prefer it works automatically without any user assistance, most color transfer methods available for this task finally fit into the scope of transferring statistical properties. [2]

Typically, images are encoded using RGB color space, in which the three channels (R, G, B respectively) are highly correlated. This implies if we want to change the appearance of a pixel's color in a coherent way, we must modify all three color channels. That complicates any color modification process and may have unpredictable results. [3] However, the color space used to manipulating the color content doesn't need to be limited to general RGB color space. By shifting, scaling and rotating the axes, we can construct different color spaces. If the three color channels of an image can be made near-independent or even independent, the image processing can take place in each of the channels independently, converting the potentially complex 3D problem to 1d transfer.

Here we will categorize the color transfer methods according to the color space they applied. At the highest level, there are two classes: parametric and non-parametric.

1.1.1 Model-based parametric approaches

Model-based methods assume parametric format of the color mapping function, classified as parametric.

Decorrelated color space

$L\alpha\beta$ color spaces and related space $CIELAB$, as well as YC_bC_r, $CIELCh^*$ and Yuv, have been adopted for many parametric transfer methods.

Based on simple statistics of global color distribution of two images, Reinhard *et al.*[3] proposed a linear transformation in decorrelated color space $L\alpha\beta$. After converting the image from RGB to $L\alpha\beta$, it transfers mean and standard deviation between each channel of the two images:

$$I_o = \frac{\sigma_r}{\sigma_t}(I_t - \mu_t) + \mu_r$$

where (μ_t, σ_t) and (μ_r, σ_r) are the mean and standard deviation of the target and reference images in $L\alpha\beta$ space. And I_t and I_o are the color distribution of the target and output images.

Correlated color space

Though considering each channel separately simplify a3D problem to three 1D problems, it cannot capture all subtleties, including local color information and interrelations. To overcome this, treating the 3D color distribution as a whole is necessary. The mapping is then reshaping the target 3D color distribution to match that of the reference.

Xiao *et al.* [4] extends Reinhard *et al.*'s work, transfering mean and covariance between images in correlated RGB space. It replaces the rotation to $L\alpha\beta$ with the axises defined by the pricipal components of each image. Formally, this can be expressed as a series of matrix transformation.

$$I_o = T_r R_r S_r S_t R_t T_t I_t$$

Here, T, R and S denote the matrics of translation, rotation and scaling derived from the target and reference images accordingly.

1.1.2 Modeless non-parametric approaches

On the contrary, non-parametric methods have no parametric format of the color transfer function and most of them directly record the mapping of the full range of color/intensity levels using a look-up table. This look-up table is usually computed from the 2D joint histogram of image feature correspondences. [5] Modeless color mapping techniques are non-parametric, also including those in correlated and decorrelated color spaces.

Decorrelated color space

Compared with simple global statistics providing only information about overall tendencies on distribution, the full distribution of values in each channel is a more faithful representation. And histogram is a useful tool to describe it.

Fecker *et al.* [9] proposed to use cumulative histogram mapping. They used a closest neighbor mapping scheme, setting the corresponding color level of the source image to each level of the target. The shape of the target histogram can be matched to that of the reference, and subsequently, the transferred image has the same color as the reference.

Correlated color space

By treating the 3D color distribution as a whole, Pitié *et al.*[6] proposed an inspiring approach called iterative color distribution transfer, matching 3D distribution through an iterative match of 1D projections. Each 1D projection of the target image is matched to that of the reference. Repeat this simple procedure for different set of axes until convergence occurs, and then the data is transformed back to the original coordinate, eventually converting the original target images into a distribution identical to the reference.

Iterative color distribution transfer can increase the graininess of the original picture, especially if the color dynamic of the two images is very different. So Pitié *et al.*[7] proposed a second stage of the method available to reduce the grain artifact through an efficient post-processing algorithm that intends to preserve the gradient field of the original target image.

1.2 Dataset generation

To enrich the diversity of our synthesized composite images, as introduced above, we choose 4 representative color transfer methods from 4 corresponding branches of color transfer approach taxonomy, global color transfer [3], global color transfer in RGB space [4], cumulative histogram matching [9] and iterative color distribution transfer with grain [7](IDT Regrain color transfer for short). And we use four splits (train2014, train2017, val2014, val2017) from the Microsoft COCO dataset [8], where the object segmentation masks are provided for each image, and 80 kinds of objects are annotated in total.

As [1] does, to transfer style to the target, proper reference image containing the object with the same semantics is selected. So for each color transfer approach and each kind of objects in COCO, we randomly choose two images containing the same sort of object, assigning one as the target image and the other as the reference. There might be several instances of the same kind object in both images. To capture better style information and generate a more obvious composite, we compare the mask areas of all instances of that object and select the largest instance region as foreground in both images. By concatenating the images and the corresponding masks together, we have foregrounds ready for applying color transfer. Combining the edited foreground with the original background of the target image generates the final composite. As such, we ensure the edited object is still realistic but not matching to the background.

To generate a large-scale dataset and more extensive variety of transferred results, we use all four color transfer methods on all 80 kinds of objects in 4 splits of COCO. Besides, as for IDT Regrain color transfer, there is one hyperparameter when reducing the grain artifacts called *smoothness*, which sets the fidelity of the original gradient field. It's higher than or equal to 0 and default by 1. When generating synthesized composite images, we use six different values of smoothness (0.01,0.1,1,10,100,1000) to enrich diversity. Thus the composites could cover different scenarios, and our network could adapt to real cases better. After the color transfer process, 79762 composite images are generated. We then also apply a released aesthetics prediction model [10] to filter out low-quality ones. We rank the prediction score and select the top 56000 images, 52000 for training and 4000 for the test.

References

[1] Tsai, Y.-H.; Shen, X.; Lin, Z. L.; Sunkavalli, K.; Lu, X.; Yang, M.-H. Deep Image Harmonization. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2017**, 2799–2807.

[2] Faridul, H. S.; Pouli, T.; Chamaret, C.; Stauder, J.; Reinhard, E.; Kuzovkin, D.; Tremeau, A. Colour Mapping: A Review of Recent Methods, Extensions and Applications. *Computer Graphics Forum 35*, 59–88.

[3] Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Computer Graphics and Applications* **2001**, *21*, 34–41.

[4] Xiao, X.; Ma, L. Color Transfer in Correlated Color Space. Proceedings of the 2006 ACM International Conference on Virtual Reality Continuum and Its Applications. New York, NY, USA, 2006; pp 305–309.

[5] Xu, W.; Mulligan, J. Performance evaluation of color correction approaches for automatic multi-view image and video stitching. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010; pp 263–270.

[6] Kokaram, A. C.; Dahyot, R. N-Dimensional Probability Density Function Transfer and its Application to Colour Transfer. In International Conference on Computer Vision (ICCV'05. 2005.

[7] Pitié, F.; Kokaram, A. C.; Dahyot, R. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding* **2007**, *107*, 123 – 137, Special issue on color image processing.

[8] Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C. L. Microsoft COCO: Common Objects in Context. Computer Vision – ECCV 2014. Cham, 2014; pp 740–755.

[9] Fecker, U.; Barkowsky, M.; Kaup, A. Histogram-Based Prefiltering for Luminance and Chrominance Compensation of Multiview Video. *IEEE Transactions on Circuits and Systems for Video Technology* **2008**, *18*, 1258–1267.

[10] Kong, S.; Shen, X.; Lin, Z.; Mech, R.; Fowlkes, C. Photo Aesthetics Ranking Network with Attributes and Content Adaptation. Computer Vision – ECCV 2016. Cham, 2016; pp 662–679.