SCIENCE CHINA
Information Sciences

• MOOP •

Situation-adaptive Neural Network for Fast Pre-computing Image Enhancement

Xinyue Li[1], Huiyu Duan[1], Jia Wang[1], Xiaohong Liu[1*], Yitong Chen[1*] & Guangtao Zhai[1]

[1]*Institute of Image Communication and Network Engineering, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai* 200240*, China*

As intelligent vision tasks become more widespread, enhancing image quality before further computational analysis is crucial. Recently, deep learning has shown potential for automated pre-computing enhancement, but it typically requires substantial computational resources and is hard to adapt to in multiple situations without re-training. In practice, image enhancement often demands flexible adjustments based on different situations and subsequent computation devices, such as optical computing. Therefore, we propose SAEnhancer, a situation-adaptive neural network for fast pre-computing image enhancement. It learns from a small sample set to achieve personalized and adaptive enhancements for various situations without re-training, using semantic-aware embedding for precise color adjustments, surpassing traditional 3D lookup tables (LUTs), and enhancing computational effectiveness in intelligent vision applications.

Introduction. With the increasing popularity of intelligent vision tasks, image enhancement in quality and color effects before further computational analysis is becoming a common practice, crucial for improving the accuracy of subsequent tasks [2]. Image enhancement, however, is inherently subjective, with situations often having diverse expectations and needs, and therefore requires flexible adjustments. For example, the computation precision of optical computing systems may be sensitive to noise in low-light or incoherent situations. Thus, enhancing images before computational tasks will be very helpful to improve the processing outcomes [1]. However unlike traditional sensors and computing devices that typically have well-investigated preferences for pre-computing image enhancement, diverse optical computing systems, or other innovative computing architectures, usually have inadequate prior information to design or train the preference, especially for unseen situations.

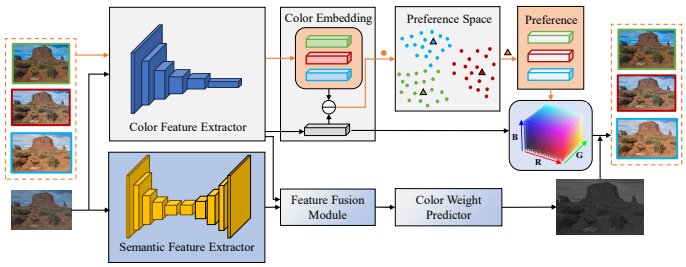
Traditional methods, typically involving manual intervention by professional retouchers or photographers, are not only time-consuming and labor-intensive but also struggle with consistency in enhancement results [2]. Recent advancements in deep learning have shown significant potential for automating and streamlining image processing, with various deep neural network (DNN) models proposed for efficient enhancement [3]. They utilize deep learning to learn color adjustment and quality improvement of paired images from large-scale datasets [4]. One type of method uses a fully convolutional network to process the image directly. The networks are trained to learn the mapping between different images [5]. These methods achieve good results but often require large memory overhead and computational cost. Another kind of methods use neural networks to learn the mapping curve or color transformation function between the downsampled image pairs. These methods usually have lower complexity [2]. LUT-based methods are commonly used in image signal processing (ISP) and image color conversion. Some recent deep learning researchers integrate lookup tables (LUTs) and neural networks and use a large number of images to train adaptive 3D LUTs for color enhancement tasks [2, 3]. The color transformation in the lookup table is always a one-to-one mapping, meaning that the color with the same pixel in the image will produce the same color transformation. However, the manual enhancement process is influenced by the image content, which cannot be imitated in the LUTs method.

To address these challenges, we propose SAEnhancer, a novel situation-adaptive lookup table method for personalized image enhancement. This method adaptively simulates situation preferences under different lighting conditions using a small set of image samples, making it suitable for enhancing images under previously unseen data and requirements. To improve traditional 3D LUTs, which rely on fixed color transformations, SAEnhancer integrates a feature fusion module and color weight predictor. These modules derive context-sensitive color weights by integrating semantic features with a color extraction network, allowing for more nuanced and context-aware adjustments.

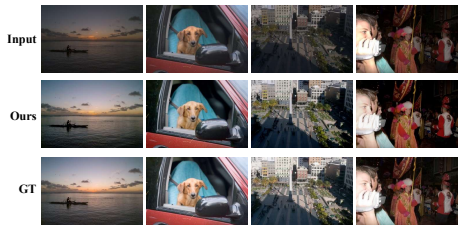
Furthermore, We use original and multiple expert-enhanced images to train a color feature extraction network to maximize the differences between various stylistic preferences. We encode these preferences into a set of style-aware weight LUTs. These LUTs represent the color preferences of different experts and can be adjusted to suit the specific aesthetic of new situations or devices, enabling personalized image enhancement that is highly adaptive and responsive to unseen situations, as shown in Figure1.

Overall Framework. We propose a situation-adaptive multi-style image enhancement method, which mainly ac-

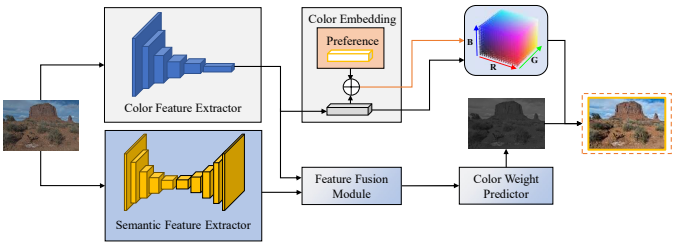
* Corresponding author (email: xiaohongliu@sjtu.edu.cn, yitongchen@sjtu.edu.cn)



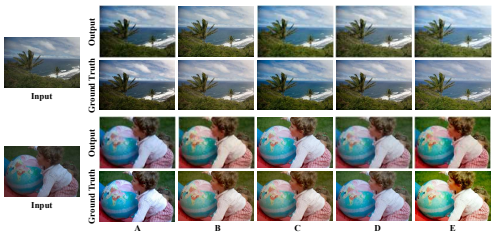
(a) Framework of the proposed SAEnhancer.



(c) Qualitative comparison of personalized image enhancement effect with the ground truth for new situation.



(b) Image enhancement of new situation preferences.



(d) Qualitative comparison of multi-situation enhancement effects with ground truth on FiveK dataset.

Figure 1 (a) Framework of the proposed SAEnhancer. (b) Image enhancement of new situation preferences. (c) Qualitative comparison of personalized image enhancement effect with the ground truth for new situations. (d) Qualitative comparison of multi-situation enhancement effects with Ground Truth on FiveK dataset

complishes two tasks: extracting color difference and fusion image semantics. The overall framework is shown in Figure 1 (a). In the first part, we use a color feature extractor to extract color embedding from the input low-quality and expert-edited images, respectively, and calculate the color difference between low-quality and expert-edited images. These color differences within the batch are then used to optimize a set of expert adjustment preference vectors that were initialized randomly. In the second part, we introduce a pre-trained lightweight segmentation network as a semantic feature extractor to extract semantic features from low-quality images. Then, we use a feature fusion module to fuse the semantic features with the color features layer by layer, resulting in a semantic-aware color feature map. We input this feature map into a color weight prediction module that generates a semantic-aware color weight map to adjust the color enhancement LUT.

Feature Fusion Module and Color Weight Predictor. The current deep learnable lookup table uses the color embedding obtained by encoding the backbone network. These adaptive LUTs can be trained by the LUTs generation module, which is made up of several fully connected layers. It is denoted as $LUT = LG(f_c)$, where f_c represents the color embedding encoded by the backbone network. For the given i-th input image x_i, we can obtain the output by applying the LUT structure transformation, i.e., $x_{out} = LUT(x_i)$.

The LUTs can only perform one-to-one color mapping. This means that pixels with the same color will be mapped to the same color. However, the expert usually chooses different color adjustment schemes based on the local content of the image. Therefore, the color adjustment process is affected by the image content. We designed the feature fusion module (FFM) to obtain image color features with semantic awareness. During the feature fusion process, we adopted a layer-by-layer fusion strategy, gradually merging the semantic and color features at different scales from shallow to deep layers of the network to bridge the heterogeneity gap. We extracted three different scales of color features, represented by ${f_{col}}^i$, and three different scales of semantic features, represented by ${f_{sem}}^i$, whose sizes are $(H/2^{4-i},\ W/2^{4-i})$, where H and W represent the length and width of the image. In the FFM module, we first obtain fusion features by applying the FFM function to both multi-scale semantic awareness f^i_{sem} and color features f^i_{col}, resulting in $f^i_{fuse} = FFM(f^i_{sem}, f^i_{col})$. To be specific, we donate the fuse block as FB. The fusion features of three scales after using fuse block are as follows: $f_0 = FB(f^0_{sem}, f^0_{col})$, $f_1 = FB(f^1_{sem}, f^1_{col})$, $f_2 = FB(f^2_{sem}, f^2_{col})$, where f_0, f_1, f_2 represent the fusion features of the three scales respectively. Then, we use the fusion blocks to fuse the fusion features of the three scales again, shown as $f_{fuse} = FB(FB(f_0, f_1), FB(f_1, f_2))$. Subsequently, we input the fusion features into the color weight predictor module (CWP) to generate the weight map with semantic awareness $w_c = cwp(f_{fuse})$. The color weight predictor is implemented by several simple convolutional layers. The obtained semantic-aware color weights are applied to the LUT output image, resulting in fine-grained color adjustment of the image.

Color Preference Extraction Module. Image enhancement is a subjective process that depends on individual preferences. Our proposed color preference extraction module uses a one-stage network to achieve multiple style enhancement. For a given input image x_i, we use a color feature extractor to obtain the high-dimensional color feature $f_{col\,i}$. Next, we map the color feature $f_{col\,i}$ to the color embedding space by the learned mapping C, which is implemented as a multi-layer perceptron (MLP). We then obtain the color embedding $c_i = C(f_{col\,i})$. We can use the color feature extractor to extract color features from the i-th image x_i and the images adjusted by experts a and b, x^a_i and x^b_i. These features can be mapped to the color space to obtain the features c_i, c^a_i, and c^b_i, respectively. Then, we subtract the color embedding of the original image from the embedding of expert a and b in the color space. This gives us the color adjustments of expert a and expert b for the input image

x_i, which are denoted as p_i^a and p_i^b, respectively. This can be formulated as $p_i^a = c_i^a - c_i = C(f_{col}{}_i^a) - C(f_{col\,i})$, and $p_i^b = c_i^b - c_i = C(f_{col}{}_i^b) - C(f_{col\,i})$.

We use the same method to calculate the differences between the images adjusted by all experts and the input images in the color space. We hope to use these differences to learn each expert's uniform preferences for images. Therefore, our goal is to make the adjustment vector of the same expert for all images as consistent as possible, and the adjustment vector of different experts for the same image as separated as possible. That is, for expert j, all features in the color adjustments of all images $p_1^j, p_2^j, ...$ are as clustered as possible, and for different experts' color adjustments $p_i^a, p_i^b, ...$ for the same image x_i are as separated as possible.

For each of the n_p experts, we define a preference vector P_i in the color space that represents their color adjustment tendency. That is, the set of preferences of all experts can be represented as $P = \{P_a, P_b, ..., P_{n_p}\}$, where n_p is the number of experts. We use the adjustment vector of each expert for each image to optimize the expert preference vector, and the optimization loss is designed as $Loss = \frac{1}{n_p}\sum_{j=0}^{n_p}\{\log(1 + \sum_{p\in P_j^+} e^{-\alpha(s(p,P_j)-\delta)}) + \log(1 + \sum_{p\in P_j^-} e^{\alpha(s(p,P_j)+\delta)})\}$, where P_j^+ represents the set of adjustment vectors of all images by the j-th expert, $P_j^+ = \sum_{i=1}^{n}\{p_i^j\}$. P_j^- represents the adjustment vector of the other experts sampled in the batch for the image, $P_j^- = \sum_{i=1}^{n}\sum_{k\neq j}\{p_i^k\}$. $s(,)$ represents the similarity between two vectors, that is, $s(p, P_j)$ represents the similarity between the expert adjustment vector p and the j-th expert preference P_j. α is the weight factor of the optimization intensity, and δ is the boundary threshold. We use the optimized expert preference vectors to fine-tune the image color according to the preferences P of each expert.

New Situation Preference. For invisible situation preference, we provide users with candidate images and color adjustment curves to collect situation preference images. We then encode these collected situation preference images by the color feature extractor. Specifically, for a new situation Y's preferred image y_i, we obtain the color feature $f_{col\,y_i} = F_{col}(y_i)$ by the color feature encoding network. We calculate the situation preference using multiple sets of color features $P_y = \frac{1}{n_s}\sum_{i=1}^{n_s} f_{col\,y_i}$, where n_s represents the total number of images provided by the user. We use the situation preference P_y to to fine-tune the image color, which can achieve style enhancement that meets situation Y's preferences.

The main structure of image enhancement of invisible situations is shown in Figure1 (b). The new situation-adjusted images and their color embeddings are marked in yellow. We first input several sets of image pairs consisting of the original image and the new situation-adjusted image into the color feature extraction network to obtain the embedding vectors of each image in the color space, and then use the mean of the differences of the color embedding of all image pairs as the new situation color adjustment preference vector. Users only need to provide a small set of preferred images to achieve better style enhancement effects.

Enhancement Performance. Our method has been rigorously tested against contemporary state-of-the-art methods on the MIT-Adobe FiveK dataset [4], with a focus on both single-situation and multi-situation image enhancement tasks. In the single-situation, our method demonstrated excellent performance, achieving a Peak Signal-to-Noise Ratio (PSNR) of 25.51 and a Structural Similarity Index Measure (SSIM) of 0.926. In the comparison of multiple situations, the proposed method improves both PSNR and SSIM substantially compared with well-accepted benchmarks, such as WB, from 17.83 to 24.50 and from 0.799 to 0.914 respectively.

New Situation Preference Enhancement. In this section, we validate the effectiveness of our method in terms of new situation enhancement. Our method can learn a new situation preference with as few as 5-10 image pairs that have not been involved in training for new situation enhancement.

As in [6], we use Lightroom to perform automatic white balance on images in the FiveK dataset and use them as a new set of adjusted images. We randomly chose 10 original images from 4500 training samples, found the corresponding 10 images from the images undergoing automatic white balance, and input those 10 image pairs into the model, which can obtain the color adjustment preferences of this situation without training. We used this situation's color adjustment preferences to adjust images from 500 testing samples, and some of the enhancement effects are shown in Figure 1 (c). We repeat the above procedure 1000 times and calculate the average PSNR is 23.13 on testing samples. It signally improves the subjective quality and quantitative indices for subsequent computing as shown in Figure 1 (d).

Conclusion and Discussion. In this work, we propose a lookup table structure guided by the situation to encode the image enhancement preference. Which achieved good performance on the FiveK datasets and proved to be effective in simulating the style of invisible situations, as shown by experimental results. It flexibly adjusts the color and quality of image enhancement based on different contexts, offering advantages in situations with limited prior knowledge, such as optical computing.

Acknowledgements The work was supported in part by the National Natural Science Foundation of China (Grant No. 62301310), the Foundation of Shanghai Jiao Tong University, and the Shanghai Pujiang Program (Grant No.22PJ1406800).

Supporting information Videos and other supplemental documents. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Chen Y, Nazhamaiti M, Xu H, Meng Y, Zhou T, Li G, Fan J, Wei Q, Wu J, Qiao F, Fang L, Dai Q. All-analog photoelectronic chip for high-speed vision tasks. Nature, 2023, 623: 48-57

2 Zeng H, Cai J, Li L, Cao Z, Zhang L. Learning Image-adaptive 3D Lookup Tables for High Performance Photo Enhancement in Real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44: 2058-2073

3 Yang C, Jin M, Jia X, Xu Y, Chen Y. AdaInt: Learning Adaptive Intervals for 3D Lookup Tables on Real-time Image Enhancement. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, 17501-17510

4 Bychkovsky V, Paris S, Chan E, Durand F. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition, 2011

5 Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, Yang J, Zhou P, Wang Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. IEEE Transactions on Image Processing, 2021, 30: 2340-2349

6 Song Y, Qian H, Du X. StarEnhancer: Learning Real-Time and Style-Aware Image Enhancement. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, 4106-4115