https://doi.org/10.2352/CIC.2025.33.1.8

An effective workflow for Colour Style Transfer

Nanlin Xu[1], Lihao Xu[2], Miaosen Zhou[1], Liangwei Chen[1] and Ming Ronnier Luo[1]*
[1]State Key Laboratory of Extreme Photonics and Intrumentaion, Zhejiang University, Hangzhou, China
[2]School of Digital Media and Art Design Hangzhou Dianzi University, Hangzhou, China
****Corresponding author: Ming Ronnier Luo, m.r.luo@zju.edu.cn***

Abstract

To perform colour rendition in digital images for different atmosphere styles is becoming important for effectively communicating visual information. While different camera brands often possess their unique feature styles, precisely reproducing and evaluating colour style effects across diverse camera systems remains significant challenge such as fast mapping and effective evaluation between the source and target styles.

To address this issue, a workflow of colour style transfer has been developed. To begin with, multi-device image database including many images was built, comprising 1550 sRGB images, for which each includes two colour charts for calibration, and proposed transfer method using a 3DLUT precisely transfers colour styles, achieving a remarkably low ΔE_{00} of 1.09 in colour charts tests. Subjective evaluation with 10 volunteers showed a perceptible small visual difference, indicating the effect of workflow achieved satisfactory performance.

To overcome the limitations of subjective testing, a Siamese network-based EfficientNet Visual Difference Evaluation Model (EVDM) was introduced, which utilized a lightweight EfficientNet, achieved Pearson correlation coefficients of 0.90 (training), 0.88 (validation), and 0.92 (overall), significantly outperforming sophisticated baseline methods based on CIEDE2000 (max 0.76). This demonstrates EVDM's superior fitting, generalization, and consistency with human perception.

1. Introduction

The colour rendition of digital images is crucial for accurately conveying visual information. Currently, major camera brands each possess their distinctive colour styles, imbuing images with unique visual characteristics.

However, two core challenges persist in image processing: the precise reproduction of these styles across heterogeneous camera systems and the objective evaluation of their effectiveness. Traditional colour transfer methods, such as polynomial and histogram matching, generally exhibit limitations in accuracy [1-3] when applied to complex natural scene, and often yielding results that diverge significantly from human subjective perception.

In recent years, deep learning approaches such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), including models like CycleGAN and StarGAN [4-5], have demonstrated powerful capabilities in image style transfer tasks. These methods have significantly improved generation quality, processing speed, and style diversity. Nevertheless, they rely heavily on large volumes of high-quality training data, a requirement often difficult to meet in practical applications.

To overcome this challenge, this study presents two primary contributions: first, we propose an innovative colour style transfer workflow based on 3DLUTs that achieves high precision without requiring massive datasets. Second, we introduce an EfficientNet Visual Difference Model (EVDM), to objectively quantify the visual differences, thereby bridging the gap between computational metrics and human perception.

2. Image Database Construction

To support the development of the proposed colour style transfer and evaluation models, a comprehensive image database encompassing a wide range of scenes, devices, and colour styles was constructed. For this database, five leading camera systems were selected: Hasselblad, Fujifilm, Nikon, Sony, and Canon, which collectively offer 31 built-in colour profiles. All colour style aberrations originating from these cameras are provided in Table 1.

Table 1: Details of capture devices and colour style profiles.

Devices	Colour style profiles
Hasselblad X2D Lens: XCD 45P	3: SD, N, PT
FUJI XT5 Lens: Sigma 18-50	10: SD, V, S, CC, NH, NS, NC, NN, E, EB
NIKON Z6 Lens: 24-70Z	6: SD, NL, VI, PT, LS, FL
SONY a7c2 Lens: Sigma 45	8: ST, PT, NT, FL, IN, SH, VV, VV2
CANON 5D2 Lens: EF 24-105	4: SD, PT, LS, ND

During the data acquisition phase, images were systematically captured across 50 diverse scenes (see Figure 1 for examples), which were meticulously selected to represent a broad spectrum of real-world environments, spanning both indoor and outdoor settings under various conditions. For each scene, all 5 camera devices and their respective 31 built-in colour styles were employed for image acquisition. To ensure consistency across captures, critical parameters—including exposure, ISO, aperture, shooting angle, and capture time—were meticulously controlled. Crucially, a PMCC (Preference Memory Colour Chart) and a MCCC (Macbeth Colour Checker Chart) were placed in every scene to ensure the accuracy of subsequent 3DLUT model construction. Ultimately, a total of 1550 high-quality images were collected in RAW format and subsequently processed into the sRGB colour space, providing a solid foundation for the entire workflow.

3. Workflow of Colour Style Transfer

An end-to-end colour style transfer workflow based on a 3D Look-Up Table (3DLUT) was deployed. The core of the workflow, illustrated in Figure 2, involves two primary stages: 3DLUT establishment and colour style transfer application.

For 3DLUT establishment, the RGB values of the target style are first converted to CIE XYZ values. Subsequently, 3DLUTs are constructed using the original style's RGB values

and these target XYZ values. For colour style transfer, when the source style's RGB values are input into the established LUT, the corresponding CIE XYZ values are obtained. These XYZ values are then converted to sRGB to derive the final RGB values, thereby completing the colour style transfer.

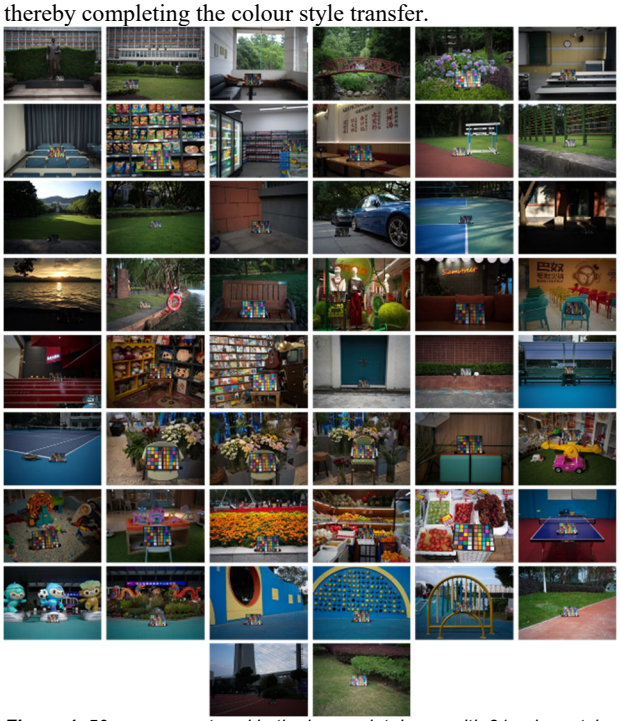

Figure 1. *50 scenes captured in the image database, with 31 colour style images for each scene.*

Specifically, the RGB2XYZ 3DLUT was generated using a lattice regression method, as reported in [6]. The term "lattice" refers to the m grid nodes $\{a_j = (R_j, G_j, B_j)\}_{j=1}^{m}$, which form a 3-dimensional rectangular grid.

The objective of lattice regression is to identify the corresponding CIEXYZ tristimulus value output vector $V_j = (X_j, Y_j, Z_j)$ that minimizes the interpolation error, i.e., the difference between the predicted XYZ values and that of the real values for the training sample datasets $\{(x_i, y_i)\}_{i=1}^{n}$ comprising n samples (where x_i is the input RGB value and y_i is the corresponding target CIEXYZ value). In our context, n is the number of colour patch samples from the MCCC and PMCC.

Each of the training samples, as well as its output values, can be calculated using the lattice nodes, as specified by Eq. (1) and (2),

$$x_i = \sum_{j=1}^{m} w_{ij}\, a_j, \quad \sum_{j=1}^{m} w_{ij} = 1 \tag{1}$$

$$\hat{y}_i = \sum_{j=1}^{m} w_{ij}\, V_j \tag{2}$$

Here, w_{ij} represents the interpolation weights determined by the position of x_i relative to the grid node a_j, through trilinear interpolation.

As a result, the optimal set of output lattice values $\hat{V} = \{V_j\}_{j=1}^{m}$ can be determined by minimizing the following objective function:

$$\hat{V} = arg\,min_V \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = arg\,min_V \sum_{i=1}^{n} (\sum_{j=1}^{m} w_{ij} V_j - y_i)^2 \tag{3}$$

Once the optimal 3DLUT grid node values $\hat{V}$ are determined, corresponding XYZ value of input RGB value can be rapidly calculated using standard trilinear interpolation (employing the weighting method from Eq. (1). Finally, these XYZ values are converted to the sRGB colour space to produce the final, stylized RGB values.

In summary, the proposed workflow, founded on a lattice regression-based 3DLUT, offers an accurate and highly efficient solution directly from a source RGB space to a target's RGB space. This streamlined approach facilitates rapid and accurate colour style transformations.

4. Workflow Performance

To validate the effectiveness of the proposed workflow, a series of objective and subjective evaluations were conducted. For these tests, the Hasselblad SD colour style was designated as the source colour style, with all 31 colour styles serving as the targets. In the practical workflow, 3DLUT colour conversion models were constructed, utilizing the colour patch data from the MCCC and

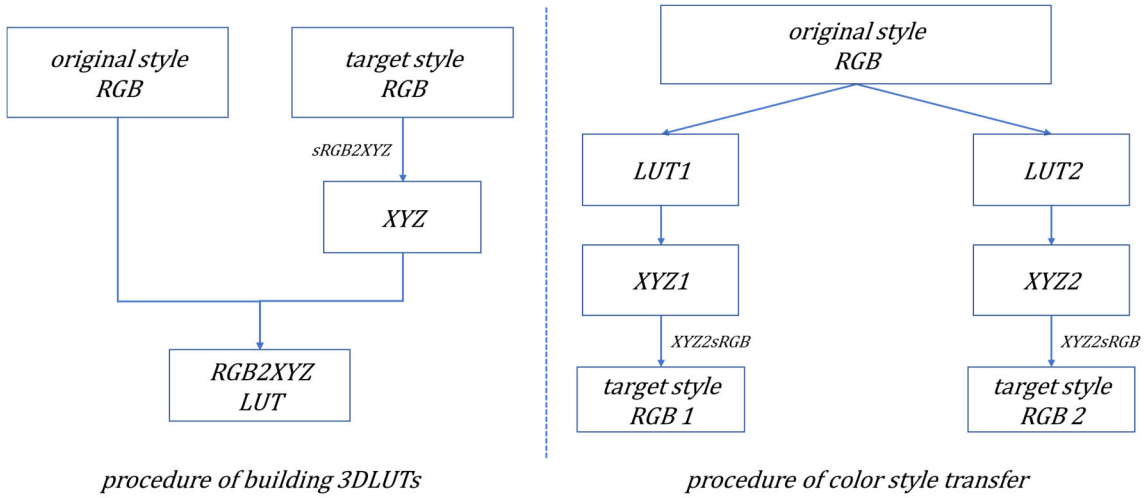

Figure 2. *Workflow of colour style transfer, comprising 3DLUT establishment and RGB transfer.*

PMCC as the training set for each scene. The grid parameter of the 3DLUT m was set as 17 to ensure optimal performance.

4.1 Objective Evaluation on Colour Charts

The performance of the workflow conversion was first evaluated objectively using the colour charts across all 50 scenes and 31 target styles. Table 2 lists the average CIEDE2000 colour differences (ΔE_{00}) [9] between the transferred colours and the target colours, using different combinations of training and testing data. As shown in the table, the mean ΔE_{00} values on the test datasets were consistently low (1.99, 1.96, and 1.09). These minimal colour differences strongly indicate that the proposed workflow provides a highly accurate foundation for colour style transfer.

Table 2: Validation performance of colour charts.

Trained on	Average ΔE_{00} with MCCC	Average ΔE_{00} with PMCC
MCCC	0.87	1.99
PMCC	1.96	0.91
2 charts	1.09	

4.2 Subjective Evaluation Experiment

To assess the workflow's performance on natural images, a subjective visual experiment was conducted. For the experiment, the 1550 source images (Hasselblad SD style) were processed to match the 31 target styles, generating the set of transferred images. Figure 7 provides a visual comparison for two of these scenes, illustrating the relationship between the source image, our workflow's result, and the ground truth target style.

The experiment was conducted using an NEC display with a resolution of 2560 × 1440 pixels. The correlated colour temperature (CCT) of the display peak white was set to 6500 K with a luminance of 300 cd/m². The Gamma coefficient was set to 2.2. The Gain-Offset-Gamma (GOG) model was used to characterize the display. The predictive accuracy of the GOG model was an average of 0.58 ΔE_{00} units over 24 Macbeth Colour Checker colours. All measurements were conducted using a Konica Minolta CS2000A tele-spectroradiometer in black surroundings. To ensure accurate colour reproduction, all experimental images were transformed into the display's native gamut using the GOG model. Observers viewed the display from a fixed distance of 60 centimeters.

Prior to the formal experiments, observers received detailed instruction, including explanations of relevant concepts and trial examples. In the experiment, two images were presented side-by-side on the display (as shown in Figure 3): one being the transferred image, and the other its corresponding target image. Observers were asked to evaluate the colour visual difference between the pair using a 6-point rating scale: 1 (no difference), 2 (just noticeable difference, JND), 3 (small difference), 4 (medium difference), 5 (large difference), and 6 (very large difference). A 60-second adaptation period to the viewing conditions was enforced at the beginning of the session. To ensure reliability, 10% of the image pairs were randomly selected for repeated evaluation. The experiment took 2.5 hours for each participant, with a break available every 50 minutes or as needed. A total of 10 observers with normal colour vision, confirmed by the Ishihara test, participated in the experiment. In total, 17,050 evaluation data points were accumulated.

Figure 3. *User interface for the subjective comparison experiment.*

4.3 Analysis of Subjective Evaluation Results

First, the reliability of the subjective data was assessed by quantifying observer consistency. Both intra-observer and inter-observer agreement were calculated using the Standardized Residual Sum of Squares (STRESS) metric as

$$\mathrm{STRESS} = 100\sqrt{\frac{\sum_{i=1}^{n}(F_s p_i - v_i)^2}{\sum_{i=1}^{n} v_i^2}} \tag{4}$$

Where p_i and v_i are the reference and batch data respectively and

$$F_s = \frac{\sum_{i=1}^{n} v_i p_i}{\sum_{i=1}^{n} p_i^2} \tag{5}$$

n is the number of image pairs. A lower STRESS value indicates better agreement.

The analysis yielded an intra-STRESS of 15.77 and an inter-STRESS of 26.59. According to established guidelines in colour science, these values are considered low, which confirms the high consistency and reliability of the collected experimental data for subsequent in-depth analysis.

The subjective experiment yielded a mean visual difference score of 2.6 across all 1550 image pairs. This score, positioned between '2: just noticeable difference (JND)' and '3: small difference' on the 6-point rating scale, validates the effectiveness of the proposed transfer workflow. It confirms that for practical applications on natural images, the framework can reproduce target colour styles, with minor and perceptually acceptable deviations.

5. Visual Difference Evaluation Model

While subjective evaluation is considered the gold standard for assessing visual effects, its high operational cost and time-consuming nature limit its large-scale application. Therefore, developing an accurate and efficient objective evaluation model is highly desired.

Traditional colour difference formulas, such as the widely used CIEDE2000 (ΔE_{00}), perform well for uniform colour patches—for instance, the mean ΔE_{00} across 2 colour charts processed by our workflow was merely 1.09. However, when applied to natural images with complex semantic content, their predictions often deviate significantly from human subjective judgments. To overcome this limitation, this study proposes the EfficientNet Visual Difference Evaluation Model (EVDM), a Siamese network designed for this task.

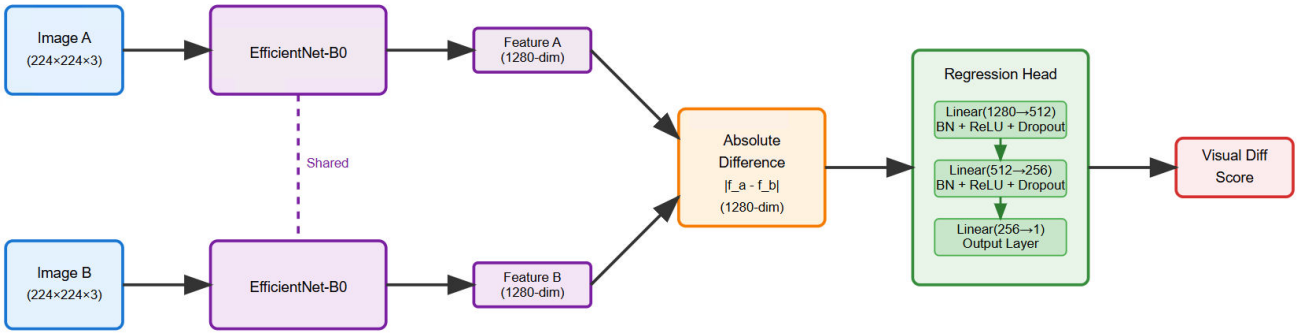

Figure 4. *Siamese network of architecture of EfficientNet visual difference model, consisting of an EfficientNet feature extractor, feature fusion module and regression head.*

5.1 The architecture of EVDM

The specific architecture of EVDM is illustrated in Figure 4. The model employs a lightweight EfficientNet-B0 as its feature extractor. EfficientNet [7] is an innovative neural network known for achieving high accuracy with fewer computational resources. It utilizes a unique "compound scaling" strategy that uniformly scales the network's depth, width, and input resolution, allowing it to outperform larger conventional networks with significantly fewer parameters.

The EVDM adopts a Siamese (shared-parameter dual-branch) architecture. In this framework, two input images were processed through identical EfficientNet feature extractors, yielding 1280-dimensional deep feature representations for each.

Subsequently, an absolute difference strategy was used to fuse these features, precisely capturing the discrepancy information between the images. The resulting fused features were then passed through a three-layer fully connected network (1280 → 512 → 256 → 1) for regression, ultimately outputting a quantized visual difference score. To enhance training stability and generalization, Batch Normalization and Dropout mechanisms were also integrated into the network.

5.2 Model Training Strategy

In the model training phase, a two-stage strategy was employed to optimize performance and effectively leverage pre-trained knowledge from the EfficientNet backbone.

Initially, the pre-trained EfficientNet backbone network was frozen, and only the regression head was trained for 10 epochs. This first stage facilitated faster initial convergence and enhanced stability by adapting the newly added regression layers to the task-specific output without disturbing the robust, generalized features learned by the backbone.

Subsequently, the entire network was unfrozen for end-to-end fine-tuning. This second stage, utilizing a relatively lower learning rate for the backbone, allowed for refined optimization and effectively prevented overfitting, ultimately leading to superior task-specific performance and improved generalization.

For optimization, the AdamW optimizer was applied. A learning rate of 1×10^{-5} was set for the backbone network and 1×10^{-3} for the prediction head, with a weight decay coefficient of 1×10^{-4}. Mean Absolute Error (MAE) was chosen as the loss function, and the ReduceLROnPlateau strategy was employed for learning rate scheduling. The dataset was split into training and validation sets in an 8:2 ratio, with a batch size of 32, for a total of 30 training epochs.

5.3 CIEDE2000 Evaluation Method Based on LightGlue

To benchmark the performance of EVDM, we developed a sophisticated baseline method using an optimized CIEDE2000 (ΔE_{00}) calculation to quantify visual differences. The process is as follows:

For each pair, the image feature matching algorithm called LightGlue [8] was employed to extract matched feature pointes. Moreover, a Homography Matrix was computed with these matched points. Based on the matrix, a perspective transformation was then applied to align the images to a common viewpoint, thereby eliminating differences in Field of View (FOV) caused by diverse systems (the extraction effect is shown in Figure 5).

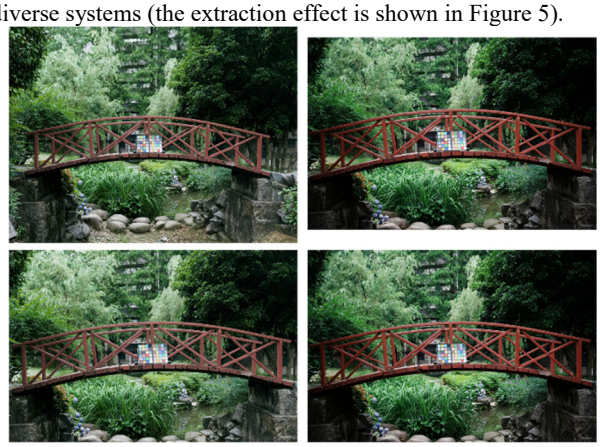

Figure 5. *Image pair comparison: pre-LightGlue algorithm vs. post-LightGlue algorithm. The top pair of images represents the transferred image and target image, while the bottom pair shows the common regions extracted.*

Next, the aligned common origin of the image pair was divided into $n\times n$ grid of sub-images. The average CIELAB values were calculated for each sub-image, and the ΔE_{00} values (given as Eq. (6)) were computed between corresponding sub-images. The final difference score was the average of all these ΔE_{00} values. To maximize the correlation with subjective ratings, the grid size n and lightness weighting factor k_L were optimized [9].

$$DE_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T\left(\frac{\Delta C'}{k_C S_C}\right)\left(\frac{\Delta H'}{k_H S_H}\right)} \quad (6)$$

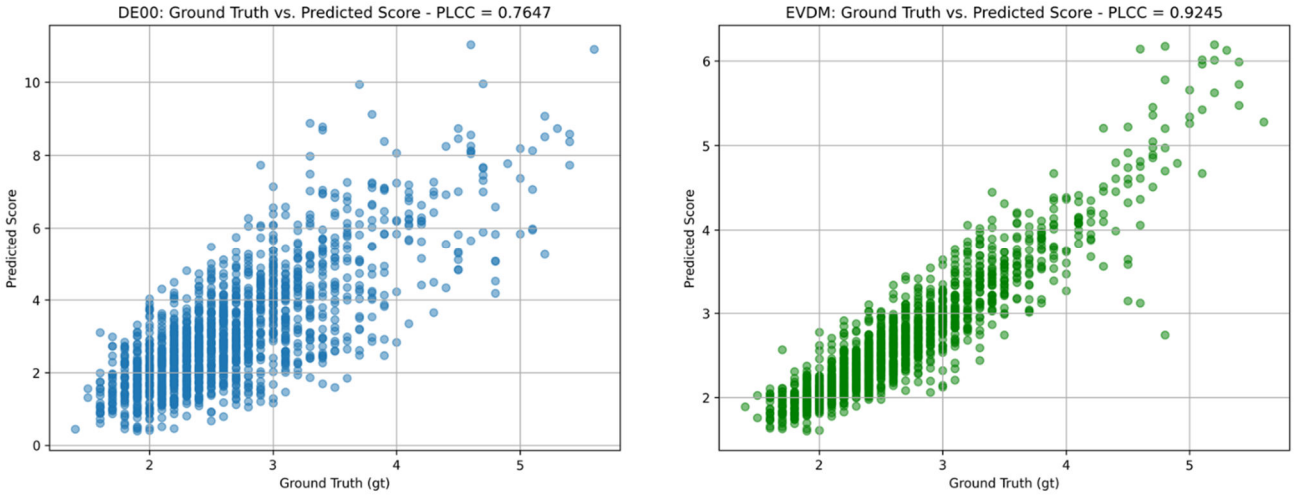

Figure 6. *Scatter plots between ground truth score and prediction score of models. The scatter plots for CIEDE2000 (ΔE_{00}) and EVDM are presented from left to right. STRSS values are 31.994 and 9.903 respectively.*

5.4 Results of visual difference model

The performance of EVDM and the baseline methods was evaluated using the Pearson Correlation Coefficient (PLCC or R) and STRESS.

EVDM demonstrated excellent and consistent performance. It achieved a R value of 0.90 on the training set, 0.88 on the validation set, and 0.92 on the overall dataset.

The proximity of the training and validation scores confirms that the model did not overfit. The high overall correlation (R=0.92) and low STRESS (9.90) validate the model's robust fitting capability and strong generalization performance.

In stark contrast, the optimized CIEDE2000 baseline, even after leveraging LightGlue and parameter tuning (optimal at n=8, k_L=3.5), yielded a maximum R of only 0.76 and a high STRESS of 31.99. Furthermore, if the LightGlue alignment step was omitted, the performance of the standard ΔE_{00} model dropped to a PLCC of 0.59 and a STRESS of 34.54. This stark difference in performance is also visualized in the scatter plots of Figure 6, where the predictions from EVDM show a much tighter correlation with subjective scores compared to the dispersed results of the ΔE_{00} methods. This highlights EVDM's superior ability to assess visual differences in complex natural images.

Table 3: Evaluation performance of 3 visual difference methods.

Method	R	STRESS
EVDM	0.92	9.90
ΔE_{00} (common)	0.76	31.99
ΔE_{00} (w/o common)	0.59	34.54

6. Conclusion

This study successfully addresses the challenges of achieving high-fidelity colour style transfer and its objective evaluation. To this end, a large-scale image database was constructed, comprising 1550 images from 5 camera brands across 50 scenes and 31 distinct colour profiles. This database served as the foundation for our two primary contributions. First, an effective workflow of colour style transfer based on 3DLUT was proposed. This method demonstrated exceptional precision, achieving a low mean colour difference (ΔE_{00} = 1.09) on 2 colour charts and a low mean visual difference score in subjective tests, confirming its ability to accurately reproduce target styles on natural images without requiring massive datasets.

The other key contribution is the development of the EVDM, a novel Siamese network to predict visual difference between images. EVDM showed a strong correlation with human perception (PLCC = 0.92, STRESS = 9.90), significantly outperforming a sophisticated baseline derived from traditional ΔE_{00} formulas. Further works will be carried out to apply neural network.

In summary, this research delivers a complete, integrated colour style transfer solution, and an efficient and reliable objective evaluation tool, offering robust technical support for achieving cross-device colour consistency and paving the way for future work in areas such as video style transfer and on-device deployment.

References

[1] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Colour transfer between images," IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 34-41, 2001.

[2] F. Pitié and A. C. Kokaram, "The spatial-colour joint histogram for video editing," IEEE Transactions on Image Processing, vol. 17, no. 5, pp. 686-699, 2008.

[3] B. Gooch and A. Gooch, Non-Photorealistic Rendering, Natick, Massachusetts: AK Peters/CRC Press, 2001.

[4] J. -Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242-2251.

[5] Y. Choi, M. Choi, M. Kim, J. -W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 8789-8797.

[6] Z. Liao and L. Xu, "Lightness modifications of the CIECAM16 and CIELAB based on the Helmholtz-Kohlrausch effect," Optics Express, vol. 32, no. 25, pp. 44918-44932, 2024.

[7] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," arXiv preprint arXiv:1905.11946, 2019.

[8] P. Lindenberger, P. -E. Sarlin, and M. Pollefeys, "LightGlue: Local Feature Matching at Light Speed," in 2023 IEEE/CVF

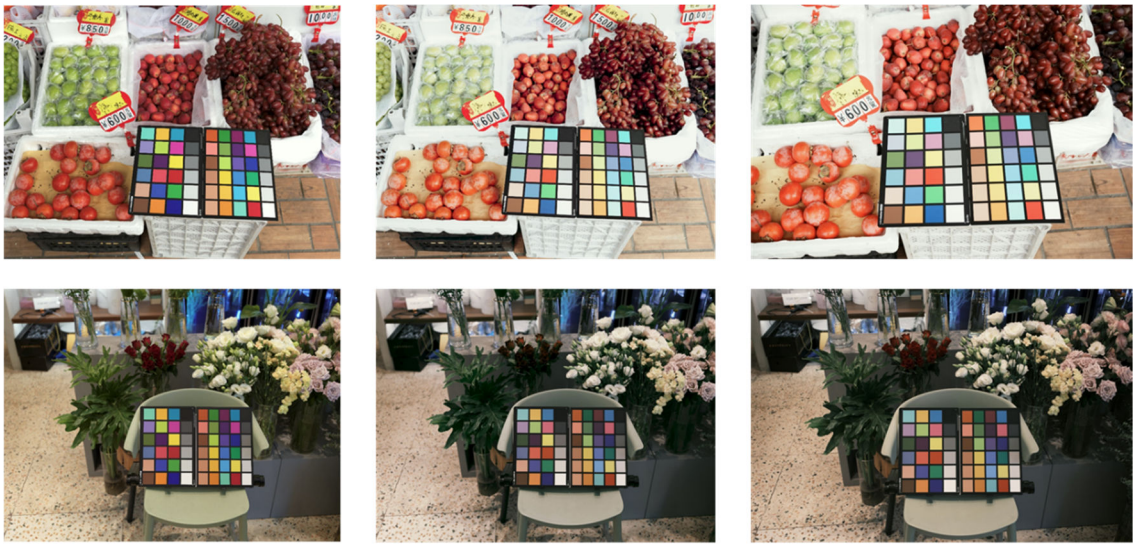

Figure 7. *Visual comparison of the proposed colour style transfer workflow for two scenes. From left to right: original image, result image generated by our workflow, and the target style image. (Top) Transfer to the Sony FL (SN-FL) style. (Bottom) Transfer to the Fujifilm NC (FUJI-NC) style.*

International Conference on Computer Vision (ICCV), Paris, France, 2023, pp. 17581-17592.

[9] S. Lee, Y. Kwak, and S. Westland, "Color Difference Evaluation for Digital Pictorial Images Using the Magnitude Estimation Method," Journal of Imaging Science & Technology, vol. 59, no. 1, pp. 10503-1–10503-8, 2015.