A Appendix / Supplemental material

A.1 Reliability Verification of Structured Weak Supervision Based on Anchor Image

To validate the feasibility of the proposed anchor-based structured weakly supervised method, we invited ten researchers specializing in image aesthetics to perform subjective rankings of color temperature aesthetics within each group of images. To reduce manual annotation costs, we first leveraged human visual perception characteristics along with the $\Delta E2000$ color difference metric. Then, using the anchor image as a reference, each group image was sampled in two directions: toward the cool shift and warm shift of color temperature. As a result, each group contained eight images (including the anchor), and the image order within each group was randomly shuffled before ranking. Participants were instructed to rank the eight images from most to least aesthetically pleasing based on their subjective perception of color temperature aesthetics. An illustration of the ranking software platform is shown in Fig. 3.

Based on anchor images, we separately aggregated the subjective ranking results for images with cool shift and warm shift, and computed the average Spearman Rank Correlation Coefficient (SRCC) and Pearson Linear Correlation Coefficient (PLCC) between the subjective rankings and the objective rankings derived from $\Delta E2000$ values. The experiment covered a total of 5,748 image groups. After removing outliers, the final dataset included 26,070 cool shift images and 14,152 warm shift images. The results show that the mean SRCC and PLCC values both exceeded 0.98, as shown in Table 5, indicating a high correlation between subjective aesthetic judgments and objective color temperature difference metrics for images with consistent color temperature shift. These results further support the validity of the proposed anchor-based ranking approach.

Table 5: SRCC and PLCC between expert rankings and objective metrics $\Delta E2000$.

Metric	Cool-shift	Warm-Shift
SRCC/PLCC	0.999/0.986	0.999/0.994

A.2 Weakly Supervised Strategy

Human vision exhibits high sensitivity to subtle shifts of color temperature, yet most IAA methods rely on holistic or multi-attribute methods, which cannot solely focus on color temperature aesthetics from confounding visual attributes. Moreover, due to the continuous spectral nature, accurately modeling the aesthetic impact of color temperature requires dense and consistent supervision. However, the Just-Noticeable-Difference (JND) nature of human perception makes continuous labeling infeasible.

To address these issues, we propose a weakly supervised strategy based on an anchor image and human visual perceptibility, as illustrated in Fig. 4. Specifically, for each group of images, based on our constructed datasets (please refer to A.3), we use the expert-adjusted image as the anchor image. By constructing the comparative relationships between images with different color temperature shifts and between the anchor image and the color temperature shift images, the model can learn to perceive aesthetic variations caused by color temperature. This method enables the construction of quasi-continuous supervision signals based on human visual perception ability, without the need for explicitly annotated continuous labels based on physical measurements.

This design offers two key advantages. First, by fixing image content and varying only the color temperature, we effectively eliminate confounding visual attributes, thereby increasing the sensitivity of the model to color temperature changes. This not only facilitates the learning of explicit aesthetic preference relations related to color temperature but also enhances the interpretability of the model, making the approach more aligned with real-world applications. Second, our method requires only a single anchor image per group to construct multiple training pairs, substantially reducing the supervision cost.

During training sample construction, we generate four types of pairwise comparisons for each reference image:

- cool-shift vs. cool-shift.
- warm-shift vs. warm-shift.
- cool-shift vs. GT.
- warm-shift vs. GT.

This strategy retains the flexibility of relative ranking while providing the model with preference signals in both directions and of varying intensity, thereby facilitating a more comprehensive understanding of aesthetic variation with respect to color temperature.

A.3 Dataset Construction

The overall data construction pipeline is illustrated in Fig. 5. We carefully selected 5,748 RAW format images from the MIT-FiveK and PPR10K datasets, with their color temperature distribution illustrated in Fig. 6. For each RAW image, we simulated the white balance adjustment mechanism, using an anchor image as the reference to generate images with varying degrees of color temperature shift. Specifically, the anchor images for the MIT-FiveK dataset were drawn from the annotated dataset by expert C, while the anchor images for the PPR10K dataset were from expert A. For textual descriptions, we employed GPT-4o to generate captions for each image encompassing the theme, content, and photography category, which were further refined through human feedback. Ultimately, we obtained 5,748 groups comprising a total of 241,416 JPEG format images. Statistical analysis of color differences in the generated images showed that the average color difference between adjacent images ($\Delta E2000$) was 0.69 with a variance of 0.74, as depicted in Fig. 7.

Test Dataset Construction. To evaluate the performance of the proposed method in assessing images with varying color temperature casts, we constructed two test datasets: a general test set (ICTAA-GP) and a portrait-specific test set (ICTAA-HP). We invited 10 researchers specializing in image aesthetics to perform pairwise preference annotations

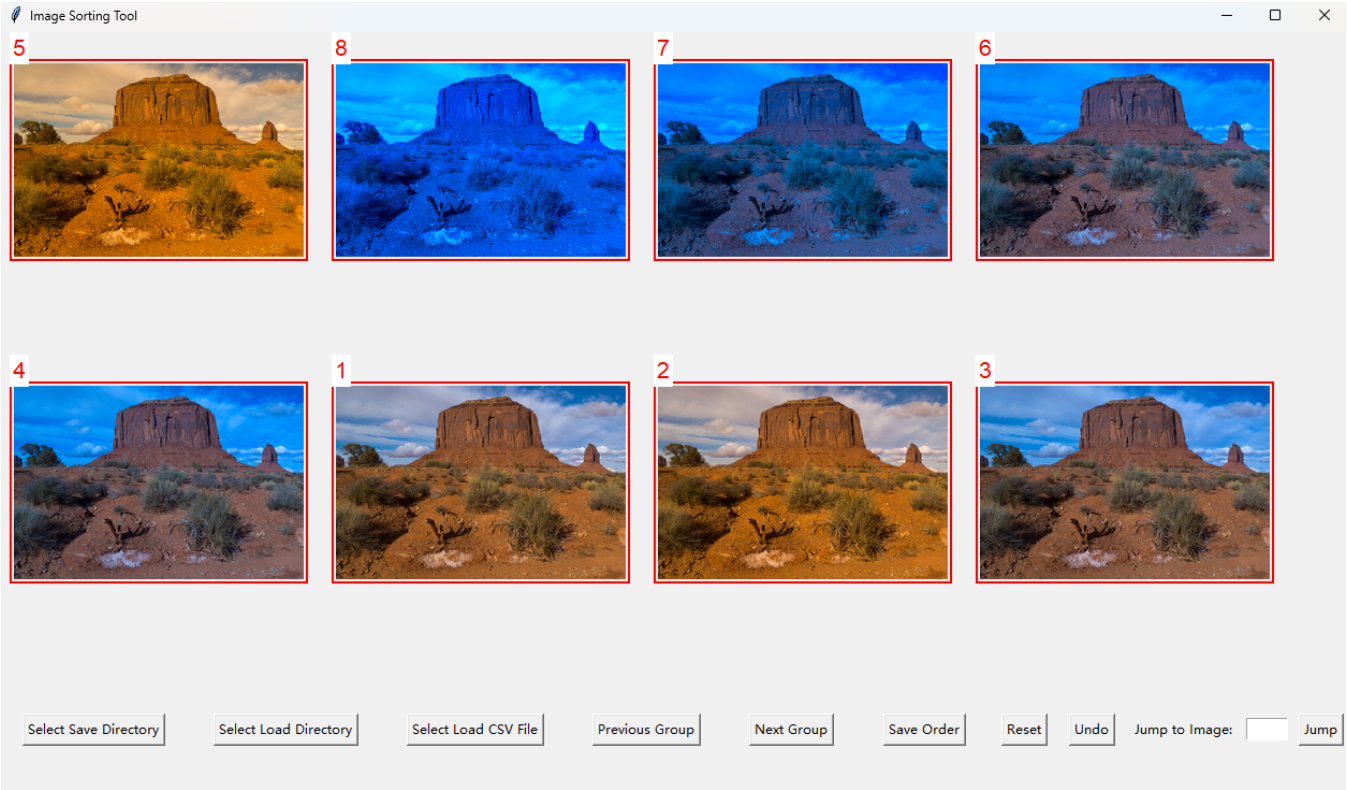

Figure 3: Ranking platform.

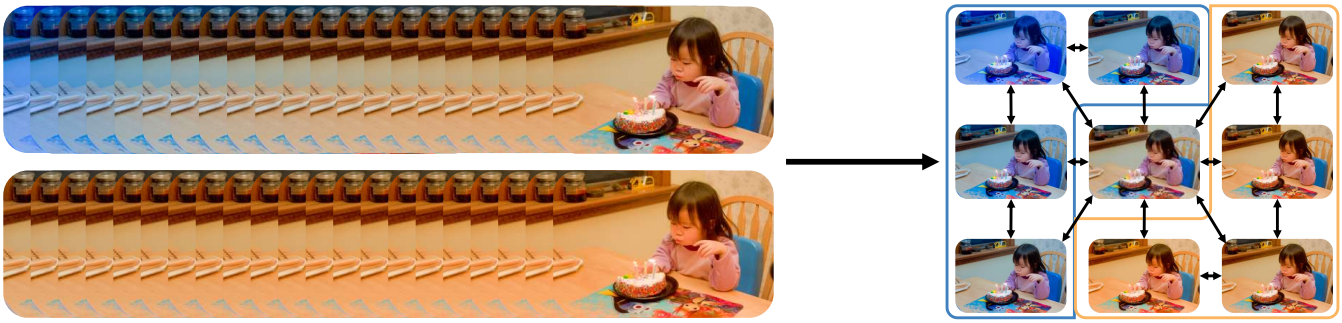

Figure 4: Weakly supervised strategy.

based on the aesthetic of color temperature. Each image pair had identical content but differed in color cast direction (cool vs. warm). Annotators rank the two images with different color temperatures according to their aesthetic preference; the annotation software platform is illustrated in Fig. 8. In total, we constructed 1,142 image pairs for ICTAA-GP and collected 1,592 pairs for ICTAA-HP.

Given that the primary concern in practical applications is whether the model can learn the optimal threshold for color temperature, we need to further evaluate whether the learned aesthetic preference trends align with the ground truth (GT) values. To this end, we constructed additional comparison samples on each of the two test datasets, by adding GT images along with their cool- and warm-shifted variants. This yielded two extended test sets: ICTAA-GF (4,628 pairs) and ICTAA-HF (6,905 pairs).

A.4 Benchmark Protocol

To compare traditional aesthetic assessment models with our proposed ICTA2Net, we designed a contrastive training protocol based on our proposed contrastive learning framework, as illustrated in Fig. 9. These conventional IAA models were incorporated into our training pipeline without any modifications to their original network architectures. For each model, the pairwise preference probability between image pairs was computed from the predicted score. We consistently used mean squared error (MSE) as the loss function across all comparison methods.

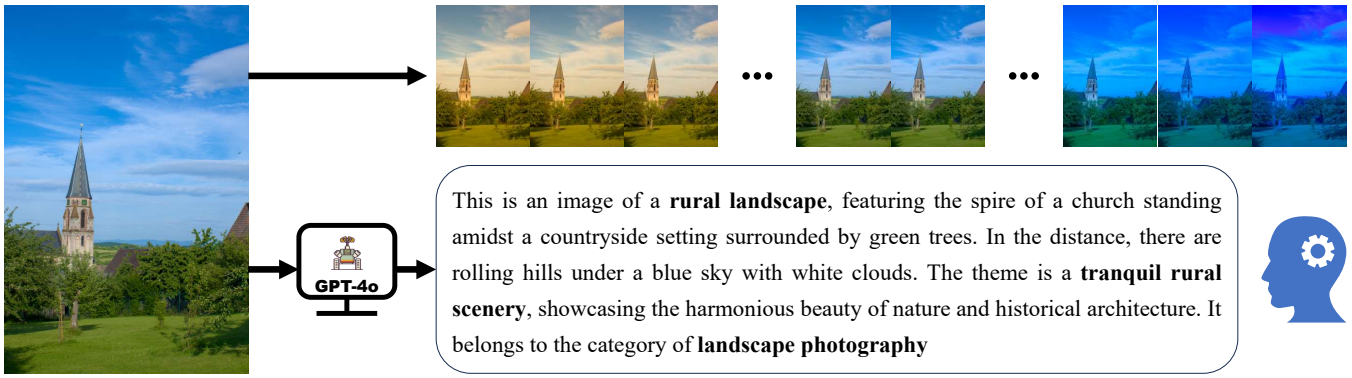

Figure 5: Dataset generation overview: for each original RAW-format image, color temperature adjustments are applied based on human visual perception characteristics to generate images with varying degrees of color temperature cast. Meanwhile, for each image, a textual description covering theme, content, and photographic category is generated using GPT-4o and refined through human feedback.

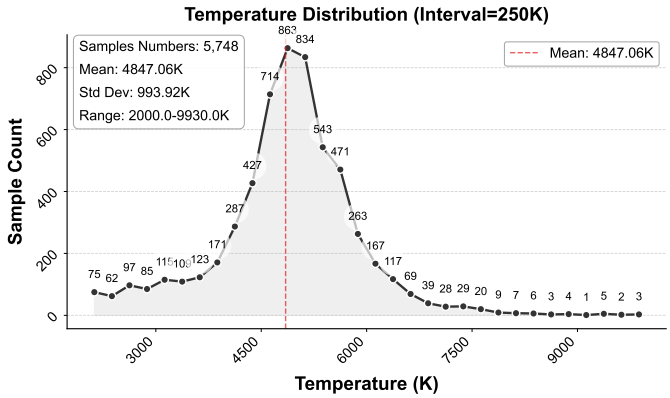

Figure 6: Color temperature distribution of RAW domain data. The overall color temperature distribution approximately follows a normal distribution and covers the range of commonly observed color temperature intervals.

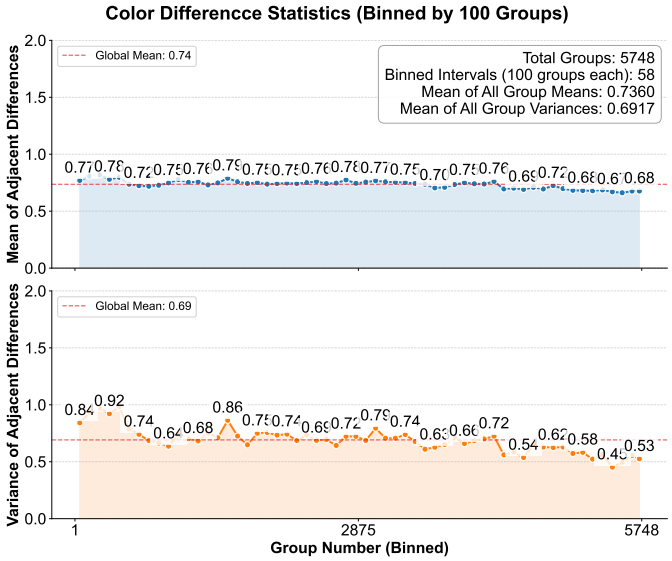

Figure 7: Adjacent color temperature difference statistic. The overall average color difference is 0.69, with a variance of 0.74.

A.5 Model Feature Output Visualization

To verify whether the model can effectively perceive both image color temperature and semantic information, we visualized the output features of ICTA2Net using t-SNE to illustrate the distribution of learned representations in the feature space, as shown in Fig. 10. Specifically, Image1 Semantic and Image2 Semantic correspond to the features F_{v1} and F_{v2} generated by the CFM module (see Fig. 2), while Color Temperature represents feature F_c extracted by the CTE. The visualization results demonstrate clear separability between color temperature and semantic features, indicating the fine-grained capability of our method to distinguish between image color temperature and contextual semantic information.

A.6 Qualitative Results

To intuitively demonstrate the ability of our model to assess image aesthetic quality related to color temperature, we randomly selected two groups of landscape and portrait images, respectively, as shown in Fig. 11, 12, 13, 14. Based on the model ranking results of model, the images are arranged from left to right and top to bottom in decreasing order of predicted aesthetic quality. It can be observed that the perceived aesthetics gradually diminishes, verifying the effectiveness of our proposed method.

A.7 Application Areas

ICTAA has broad application prospects. In this paper, we envision two potential use cases.

Photography Guidance. Color temperature aesthetic assessment can provide photographers with scientifically grounded and quantifiable guidance for adjusting color temperature, helping them select parameters that better align

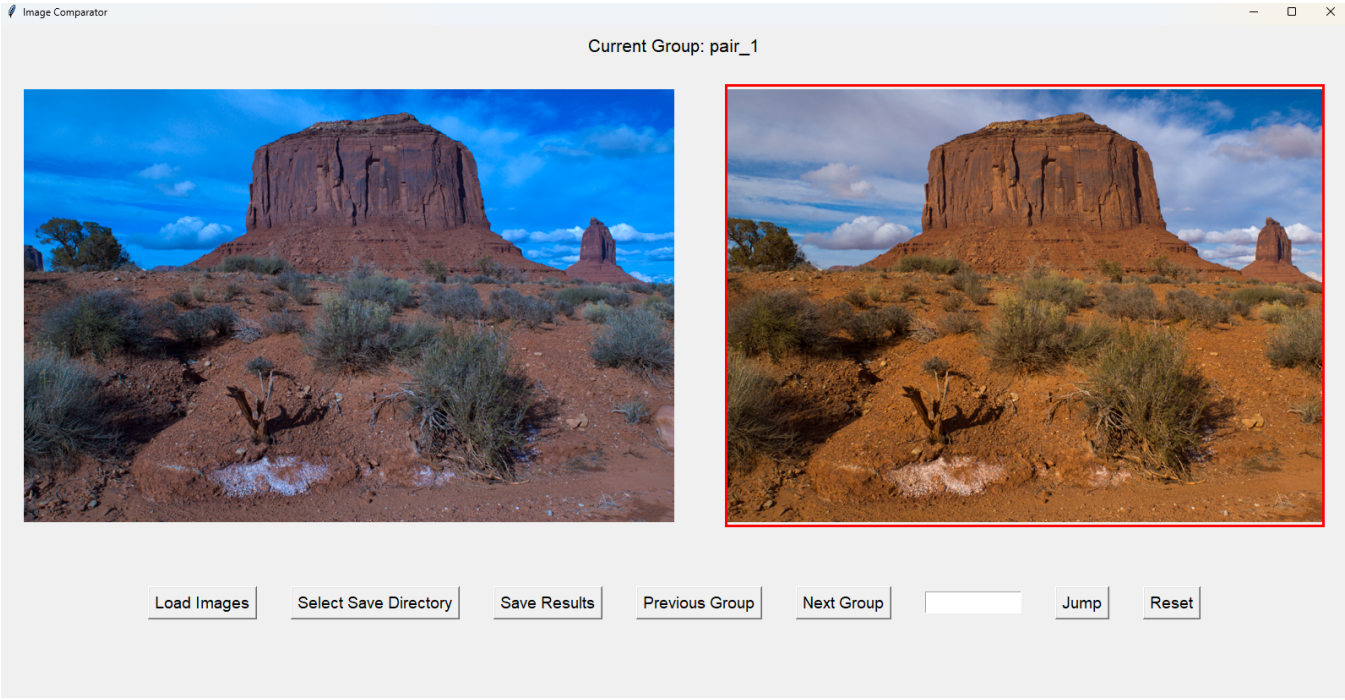

Figure 8: Scoring platform for the test dataset.

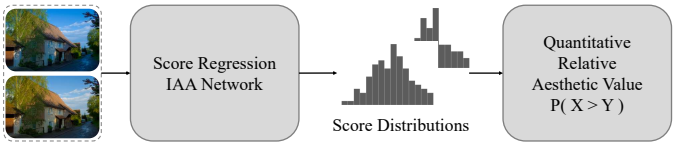

Figure 9: Contrastive training pipeline for traditional regression-based IAA models. This pipeline is designed to perform contrastive training on existing IAA methods, enabling a comparative evaluation with our proposed ICTA2Net approach.

with visual aesthetics during shooting. By analyzing the aesthetic quality of color temperature in real time, it effectively helps avoid issues such as color distortion or insufficient atmospheric depth caused by improper settings. As a result, it enhances the artistic expressiveness and perceptual consistency of photographic works.

Film Post-production Color Grading. In the film industry, color temperature adjustment not only affects the realism of on-screen colors but also plays a crucial role in conveying emotions and establishing narrative atmosphere. Color temperature aesthetic assessment can provide colorists with objective and fine-grained aesthetic feedback, assisting in the development of color grading schemes that better align with the storyline and viewers' visual preferences.

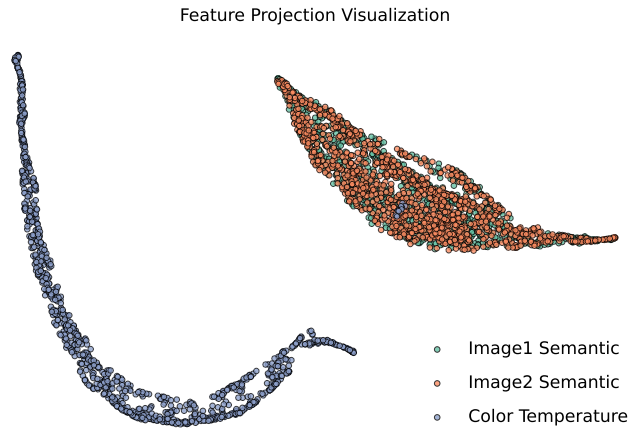

Figure 10: t-SNE visualization of the model's output features. Image1 Semantic and Image2 Semantic represent the semantic features of the two input images, respectively, while Color Temperature denotes the extracted color temperature features.

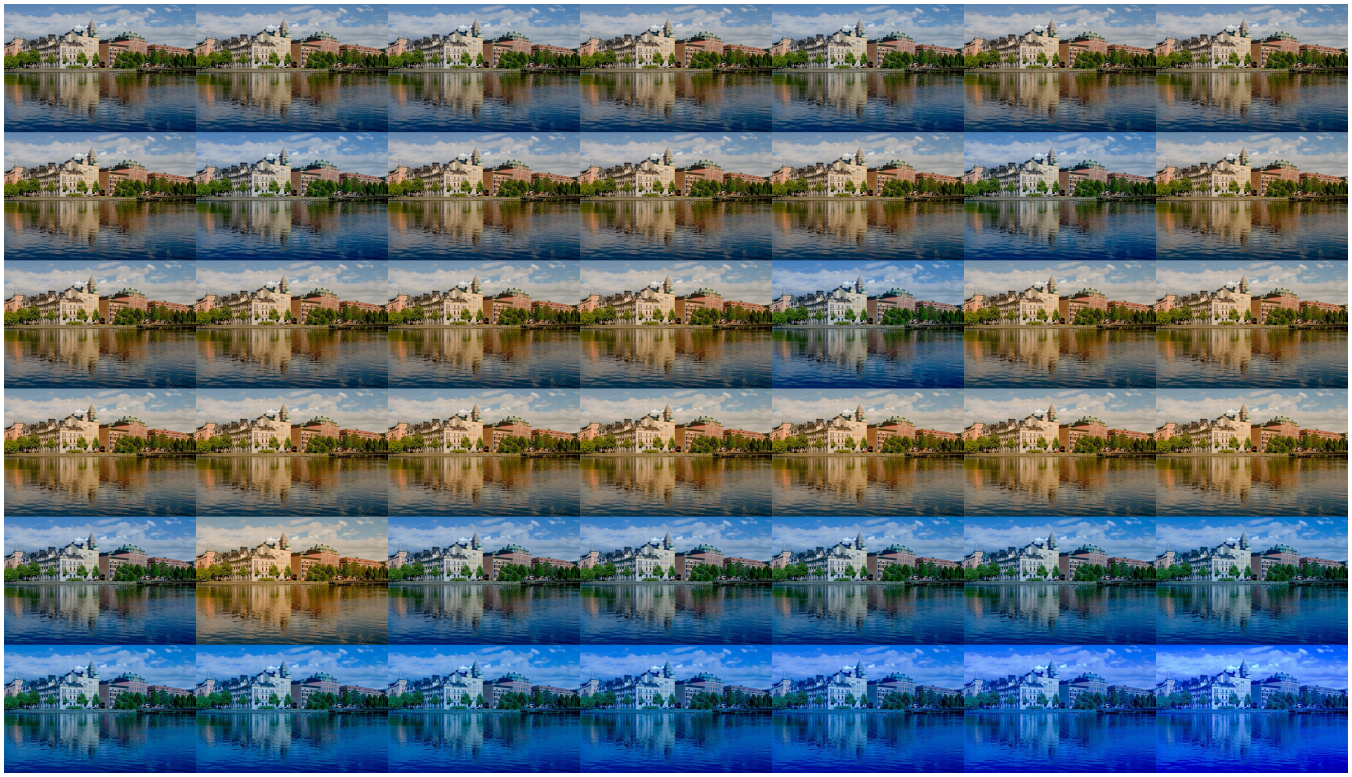

Figure 11: Visualization of model ranking results: aesthetic scores decrease progressively from left to right and top to bottom.

Figure 12: Visualization of model ranking results: aesthetic scores decrease progressively from left to right and top to bottom.

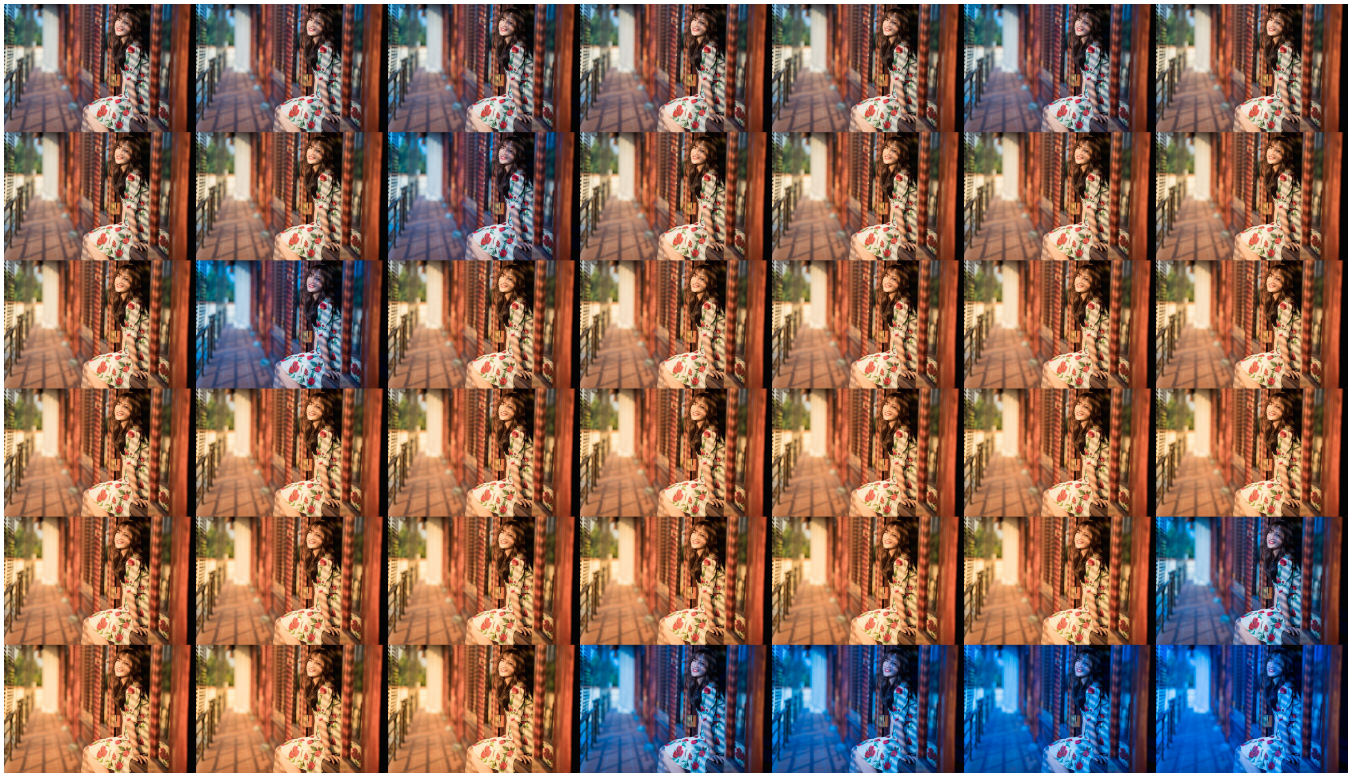

Figure 13: Visualization of model ranking results: aesthetic scores decrease progressively from left to right and top to bottom.

Figure 14: Visualization of model ranking results: aesthetic scores decrease progressively from left to right and top to bottom.