Scene-Aware Image Aesthetic Quality Assessment

Maedeh Daryanavard[1][0009-0009-3892-7016], Asadollah Shahbahrami[2][0000-0002-5195-1688], Reza Hassanpour[1][0000-0001-8649-9671], and Georgi Gaydadjiev[3][0000-0002-3678-7007]

[1] Faculty of Science and Engineering, University of Groningen, The Netherlands
{m.daryanavard, r.zare.hassanpour}@rug.nl,
[2] Department of Computer Engineering, University of Guilan, Iran
shahbahrami@guilan.ac.ir,
[3] Department of Quantum and Computer Engineering, Delft University of Technology, The Netherlands
G.N.gaydadjiev@tudelft.nl

Abstract. Image aesthetic attribute assessment provides explainable outputs for Image Aesthetic Quality Assessment (IAQA), evaluating attributes such as rule of thirds, symmetry, and lighting. These attributes are context-dependent, as their importance varies across different photography scenes. However, most attribute-based IAQA methods remain scene-agnostic, limiting their ability to model scene–attribute dependencies. We propose a scene-aware IAQA model based on a vision transformer that extracts multi-level features and integrates learned scene embeddings within a two-tower module to capture both general and scene-specific patterns, with adaptive gating for context-aware fusion. Experimental results show improved correlation for both overall score and attribute prediction, outperforming state-of-the-art attribute-based IAQA methods.

Keywords: Image Aesthetic, Image Aesthetic Quality Assessment, Scene-Aware Aesthetic Attributes

1 Introduction

Image Aesthetic Quality Assessment (IAQA) aims to predict human perception of image aesthetics. Due to the subjective nature of aesthetics, deep learning methods are widely adopted. The IAQA typically involves four tasks: scoring, distribution, attribute, and description. Scoring estimates aesthetic via classification or regression, while distribution models user preference variability. Attribute evaluation focuses on visual factors, such as color, lighting, and composition. Description generation explains aesthetic appeal using image content or user feedback. Most existing approaches focus on aesthetic scoring, with only a few addressing attribute assessment jointly [3]. Moreover, aesthetic attributes are scene-dependent; for instance, depth of field is important in nature photography, while it is less significant in urban or architectural scenes [4]. Existing approaches

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

often ignore scene categories, limiting their ability to model attribute–score relationships and reducing prediction accuracy. We hypothesize that scene information improves IAQA through adaptive modeling. However, the widely used Aesthetic Visual Analysis Dataset (AADB) [12] lacks scene annotations. We therefore extend AADB with scene labels using the Composition Assessment Database (CADB) [26], a subset of AADB annotated with attributes and scene categories. Scene labels are transferred for overlapping images and predicted for the rest, resulting in a scene-annotated AADB for training a scene-aware IAQA model. The main contributions are as follows:

- An annotated scene AADB dataset is constructed by transferring scene labels from CADB and predicting missing annotations using a ResNet-18 classifier.
- A scene-aware IAQA approach is proposed that leverages scene information to adapt attribute importance across content types.

2 Related Work

Traditional attribute-based IAQA relied on handcrafted features [10, 20], which poorly capture high-level semantics. Deep learning methods address these limitations by modeling multi-level features. Moreover, aesthetic attributes and overall score are influenced by visual context. For example, the scenes Architecture and Cityscape cluster closely based on mutual information, indicating similar relationships between attributes and overall score [4]. Existing IAQA methods follow two paradigms: attribute-based approaches and scoring-based approaches using datasets such as AVA [22]. In this work, we focus on attribute-based IAQA. Early work introduced the AADB dataset and joint attribute–score modeling [12]. Subsequent methods explored attribute-based IAQA using the AADB [2, 6, 14, 24, 19, 17]. Other methods combine attributes with theme features using external datasets [15, 7]. Several methods predict overall score using attributes [23, 25, 1, 9, 8, 18, 5, 13], while multi-modal approaches incorporate textual guidance [16, 11]. Despite these advances, most methods remain scene-agnostic due to lack of annotations, rely on external datasets that introduce bias, or focus on only a subset of attributes or the overall score. In contrast, we propose a scene-aware IAQA framework that integrates attributes and scene categories in a unified dataset to improve both score and attribute prediction.

3 Proposed Approach

Image aesthetic quality assessment depends on low-level features, global composition, spatial relationships, and semantic context. As shown in Fig. 1, the model has three stages motivated by the hierarchical nature of aesthetic perception: (i) multi-level feature extraction, (ii) a scene-aware two-tower prediction module that captures both general and scene-specialized aesthetic patterns, and (iii) a fusion gate that adaptively combines the predictions from the two towers.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

3.1 Multi-level Feature Extraction

We adopt a ViT backbone as the main feature extractor and extract intermediate representations from multiple transformer blocks to capture hierarchical visual information. We tap three blocks with indices $b_1 = 2$, $b_2 = \lfloor L/2 \rfloor$, and $b_3 = L - 1$ using zero-based indexing, where L is the total number of blocks. Let $\mathbf{Z}_{b_i} \in \mathbb{R}^{B\times(N+1)\times D}$ denote the output token sequence of block b_i, where B, N, and D denote batch size, patch count, and feature dimension. The CLS token is denoted by $\mathbf{z}_{b_i}^{\text{CLS}} \in \mathbb{R}^{B\times D}$ and the j-th patch token is denoted by $\mathbf{z}_{b_i,j}^{\text{patch}} \in \mathbb{R}^{B\times D}$. A pooled feature vector $\mathbf{p}_{b_i} \in \mathbb{R}^{B\times 2D}$ is obtained by concatenating the CLS token with the mean pooled patch tokens as follows:

$$\mathbf{p}_{b_i} = \Big[\frac{1}{N}\sum_{j=1}^{N} \mathbf{z}_{b_i,j}^{\text{patch}} \;;\; \mathbf{z}_{b_i}^{\text{CLS}}\Big] \in \mathbb{R}^{B\times 2D}. \tag{1}$$

The multi-level visual representation is then formed by concatenating pooled features from all selected blocks:

$$\mathbf{h} = \big[\mathbf{p}_{b_1}; \mathbf{p}_{b_2}; \mathbf{p}_{b_3}\big] \in \mathbb{R}^{B\times 6D}. \tag{2}$$

3.2 Scene-Aware Two-Tower Prediction

To incorporate high-level scene semantics, each image is assigned a scene label $c \in \{0, \ldots, 9\}$. We map c to a learnable scene embedding $\mathbf{s}_c \in \mathbb{R}^{d_s}$, with $d_s = 128$. For a batch, this produces $\mathbf{S} \in \mathbb{R}^{B\times d_s}$. The final representation is obtained by concatenating visual features with the scene embedding:

$$\tilde{\mathbf{h}} = \big[\mathbf{h}; \mathbf{s}_c\big] \in \mathbb{R}^{B\times(6D+d_s)}. \tag{3}$$

Since aesthetic perception varies across scene categories, a two-tower module Column A and Column B is designed to capture both general and scene-specialized aesthetic patterns. For each image, the scene label retrieves a learned embedding, concatenated with pooled ViT features and used to condition the module via scene-aware gating A_{gate} and F_{gate} for adaptive expert selection and fusion.

Specialized Experts: Column A consists of three expert heads: one generic and two scene-specialized experts for scene categories 3 (Human) and 9 (Static), selected based on attribute relationship clustering [4]. We focus on the two scenes with the strongest attribute–score divergence to avoid over-parameterization and data sparsity. Each expert operates on the scene-conditioned representation $\tilde{\mathbf{h}}$. A scene-aware gating network A_{gate}, implemented as a two-layer multilayer perceptron, maps the scene embedding $\mathbf{s}_c$ to logits, which are normalized via softmax to obtain weights $\mathbf{w}^A$. To regulate the participation of scene-specialized experts, a hard routing mechanism is applied. For samples with scene label $c \notin \{3, 9\}$, the corresponding experts are deactivated by zeroing their gating

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

weights, followed by renormalization. The raw predictions from Column A are computed as weighted combinations of the expert outputs:

$$\hat{y}^A_{\text{raw}} = w^A_g \hat{y}^A_g + w^A_3 \hat{y}^A_3 + w^A_9 \hat{y}^A_9, \tag{4}$$

$$\hat{\mathbf{a}}^A_{\text{raw}} = w^A_g \hat{\mathbf{a}}^A_g + w^A_3 \hat{\mathbf{a}}^A_3 + w^A_9 \hat{\mathbf{a}}^A_9. \tag{5}$$

For scenes 3 and 9, the corresponding scene-specialized expert is allowed to contribute through the gating mechanism. For other scenes, these experts are suppressed via the gating mask, resulting in zero contribution and routing the prediction through the generic expert. The final output of Column A is an overall score and a full set of attribute predictions.

General Predictor: Column B provides a general prediction branch that estimates aesthetic score and attributes using a multilayer perceptron $f_B(\tilde{\mathbf{h}})$. This branch is trained on all scene categories and serves as a robust alternative when scene labels are noisy or scene-specialized experts in Column A are unreliable. The outputs are passed through a configuration-dependent activation function, while dimension-wise activations are applied to attribute predictions according to the annotation scale.

3.3 Fusion Gate

The predictions from Columns A and B are adaptively combined using a second scene-aware gating module $F_{\text{gate}}(\cdot)$. The fusion gate F_{gate} maps $\mathbf{s}_c$ to fusion weights $\mathbf{w}^F = [w^F_A, w^F_B]$ via softmax. The final fused predictions are computed as weighted combinations of the column outputs:

$$\hat{y}_{\text{fuse}} = w^F_A \hat{y}^A + w^F_B \hat{y}^B, \tag{6}$$

$$\hat{\mathbf{a}}_{\text{fuse}} = w^F_A \hat{\mathbf{a}}^A + w^F_B \hat{\mathbf{a}}^B. \tag{7}$$

The scene-aware model is illustrated in Figure 1. The ViT backbone is tapped in the early, middle, and late blocks. From each tap, the mean of tokens is pooled and concatenated. A learned scene embedding is appended and fed to two gates: A_{gate} from Column A, which includes one generic expert and two scene-specialized experts, and F_{gate}, which fuses Column A with general Column B. The outputs are an overall aesthetic score and 11 aesthetic attributes.

4 Experiments and Analysis

4.1 Dataset

We use AADB [12] and CADB [26]. The AADB dataset contains approximately 10,000 images and is well-suited for attribute-based IAQA, providing an overall score and 11 attributes. However, it lacks scene labels; therefore, scene annotations are incorporated from the CADB dataset. The CADB contains 9,497 from

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

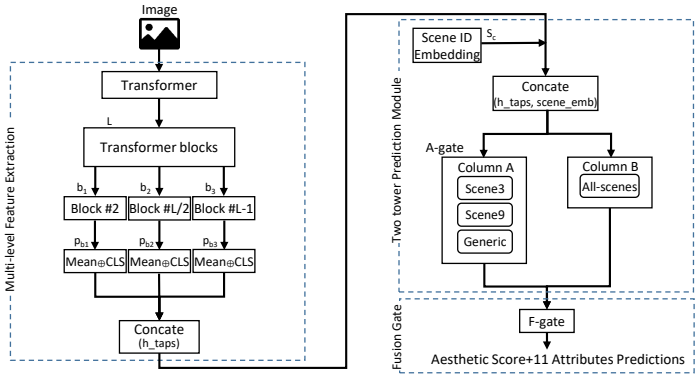

Fig. 1. Scene-aware IAQA architecture. Multi-level ViT features and scene embeddings are processed by a two-tower module and fused to predict overall score and 11 attributes.

AADB dataset. Importantly, CADB provides both attributes and scene category annotations, making it suitable for our task. Scene categories are divided into ten classes including animal, architecture, cityscape, human, indoor, landscape, night, other, plant, and static. To assign scene labels to AADB, overlapping images with CADB were matched and their annotations transferred, covering most samples. The remaining 461 images were labeled using a ResNet-18 classifier trained on CADB, which was split into 85.8% training, 9.5% validation, and 4.6% test sets, achieving 77.66% validation accuracy.

4.2 Implementation Details

The model is implemented in PyTorch with a small DINOv2-initialized ViT backbone [21]. The backbone comprises 12 transformer blocks, with features extracted from blocks 2, 6, and 11 to capture multi-level representations; for each block, the CLS token is concatenated with mean-pooled patch tokens. These features are fused with a 128-dimensional scene embedding before being fed to the prediction heads. Both Column A and Column B use SiLU activations, while the gating modules (A-gate and F-gate) followed by softmax to produce mixture weights. A sigmoid is used for score prediction, and attribute outputs are activated according to their ranges (sigmoid for $[0, 1]$, tanh for $[-1, 1]$). Images are resized to 336×336, square-padded, and augmented with random flips, affine transforms, and color jitter. The model is trained for 25 epochs using AdamW (batch size 4). Validation and testing use resizing, normalization, and test-time horizontal flip.

4.3 Performance Evaluation

We compare the proposed model with state-of-the-art methods on AADB using Spearman's rank-order correlation coefficient (SRCC). As shown in Table 1, the

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

Table 1. The SRCC comparison for aesthetic attributes and overall score on AADB dataset. Columns include the backbone, Balancing Elements, Color Harmony, Interesting Content (Content), Depth of Field (DoF), Lighting (Light), Motion Blur (Motion), Object Emphasis (Object), Rule of Thirds (RoT), Vivid Color (Vivid), Repetition, Symmetry, and the overall aesthetic score (Score).

Ref.	Backbone	Balancing Element	Color Harmony	Content	DoF	Light	Motion	Object	RoT	Vivid	Repetition	Symmetry	Score
[12]	AlexNet	0.220	0.471	0.508	0.479	0.443	–	0.602	0.225	0.648	–	–	0.678
[19]	ResNet50	0.186	0.475	0.584	0.495	0.399	–	0.666	0.178	0.681	–	–	0.689
[6]	EfficientNet-B4	0.331	0.517	0.599	0.677	0.515	–	0.677	0.273	0.706	–	–	0.706
[2]	ResNet-50	0.339	0.539	0.580	0.553	0.488	0.416	0.673	0.279	0.715	0.530	0.475	0.708
[14]	ResNet-50	0.291	0.511	0.588	0.566	0.487	–	0.658	0.264	0.722	–	–	0.737
[24]	VGG16	0.267	0.484	0.593	0.497	0.445	0.109	0.639	0.235	0.669	0.355	0.177	0.707
Ours	Vision Transformer	**0.347**	0.493	**0.653**	0.547	**0.515**	0.182	**0.719**	**0.309**	0.694	0.419	0.271	**0.743** ± **0.03**

Table 2. Ablation study of the proposed model (SRCC per attribute and overall score).

Model	Towers	Scene Embedding	Hard Routing	Balancing Element	Color Harmony	Content	DoF	Light	Motion	Object	RoT	Vivid	Repetition	Symmetry	Score
Baseline	1	×	×	0.223	0.434	0.578	0.514	0.460	0.186	0.643	0.237	0.599	0.422	0.200	0.683
Single Tower General	1	✓	×	0.280	0.504	0.624	0.539	0.495	0.152	0.671	0.292	0.665	0.412	0.222	0.712
Full No-Hard-Routing	2	✓	×	0.316	0.492	0.618	0.531	0.485	0.161	0.693	0.281	0.663	0.392	0.248	0.706
Full Model	2	✓	✓	**0.347**	0.493	**0.653**	**0.547**	**0.515**	0.182	**0.719**	**0.309**	**0.694**	0.419	**0.271**	**0.743**

proposed model achieves the highest SRCC of 0.743 and the best results on five attributes, including Balancing Element, Content, Light, Object, and Rule of Thirds. These results highlight the effectiveness of scene-aware modeling for both overall and attribute-level prediction. Per-attribute analysis shows greater improvements for semantics and composition-related attributes (e.g., *Interesting Content*, *Object Emphasis*, *Balancing Elements*, *Rule of Thirds*), while gains are smaller for attributes such as *Color Harmony*, *Depth of Field*, and *Vivid Color*, which rely more on low-level visual properties. The model also achieves a Pearson Linear Correlation Coefficient (PLCC) of 0.748 on the test set. Although PLCC was not reported in previous studies, our result further confirms a strong correlation with ground-truth. Bootstrap resampling (10,000 iterations) yields 95% confidence intervals of [0.715, 0.776] for SRCC and [0.722, 0.775] for PLCC, indicating stable performance with an uncertainty of approximately ±0.03.

4.4 Ablation Study

Four variants are evaluated: *Baseline*, *Single Tower General*, *Full Model No-Hard-Routing*, and *Full Model*. The Baseline uses a single ViT tower without scene information, while Single Tower General adds scene embeddings. Full model No-Hard-Routing introduces a dual-tower with soft gating, and the Full Model further incorporates scene-aware hard routing. Table 2 reports SRCC for all attributes and the overall score. Results show scene information improves performance, with Single Tower General increasing SRCC from 0.683 to 0.712. Scene context particularly benefits attributes such as *Balancing Elements*, *Color Harmony*, *Content*, and *RoT*. The Full No-Hard-Routing model does not consistently outperform the single-tower, as fully learned routing increases complexity and can lead to unstable expert assignment. Full Model achieves the best performance (SRCC = 0.743). Specialized experts focus on scene-relevant attributes, while shared experts maintain generalization, improving predictions.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

5 Conclusions

In this paper, we propose a scene-aware IAQA model that integrates scene information into attribute assessment. By extending AADB with scene annotations, the model leverages multi-level ViT features, scene embeddings, and adaptive gating to improve both attribute and overall score prediction. Experimental results show improved performance over state-of-the-art methods. Future work will focus on improving interpretability through richer attribute–scene annotations and developing dynamic scene–attribute modeling and visualization techniques.

References

1. Celona, L., Leonardi, M., Napoletano, P., Rozza, A.: Composition and style attributes guided image aesthetic assessment. IEEE Transactions on Image Processing **31**, 5009–5024 (2022). https://doi.org/10.1109/TIP.2022.3191853
2. Chen, Z.: Data covariance learning in aesthetic attributes assessment. Journal of Applied Mathematics and Physics **08**, 2869–2879 (2020). https://doi.org/10.4236/jamp.2020.812212
3. Daryanavard C., M., Shahbahrami, A., Hassanpour, R., Gaydadjiev, G.: Deep learning based image aesthetic quality assessment- a review. ACM Comput. Surv. **57**(7) (2025). https://doi.org/10.1145/3716820
4. Daryanavard C., M., Shahbahrami, A., Hassanpour, R., Gaydadjiev, G.: Do scenes matter? analyzing scene-aware attribute learning for image aesthetics. In: Proceedings of the 11th International Congress on Information and Communication Technology and Excellence Awards (2026), to appear
5. Duan, J., Chen, P., Li, L., Wu, J., Shi, G.: Semantic attribute guided image aesthetics assessment. In: IEEE International Conference on Visual Communications and Image Processing. pp. 1–5 (2022). https://doi.org/10.1109/VCIP56404.2022.10008896
6. Gajjala, V., Mukherjee, S., Thakur, M.: Measuring photography aesthetics with deep cnns. IET Image Processing **14**, 1561–1570 (2020). https://doi.org/10.1049/iet-ipr.2019.1300
7. Huang, Y., Li, L., Chen, P., Wu, J., Yang, Y., Li, Y., Shi, G.: Coarse-to-fine image aesthetics assessment with dynamic attribute selection. IEEE Transactions on Multimedia **26**, 9316–9329 (2024). https://doi.org/10.1109/TMM.2024.3389452
8. Jin, X., Li, X., Lou, H., Fan, C., Deng, Q., Xiao, C., Cui, S., Singh, A.K.: Aesthetic attribute assessment of images numerically on mixed multi-attribute datasets. ACM Trans. Multimedia Comput. Commun. Appl. **18**(3) (2023). https://doi.org/10.1145/3547144, https://doi.org/10.1145/3547144
9. Jin, X., Lou, H., Huang, H., Li, X., Li, X., Cui, S., Zhang, X., Li, X.: Pseudo-labeling and meta reweighting learning for image aesthetic quality assessment. IEEE Transactions on Intelligent Transportation Systems **23**(12), 25226–25235 (2022). https://doi.org/10.1109/TITS.2022.3207152
10. Ke, Y., Tang, X., Jing, F.: The design of high-level features for photo quality assessment. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 419–426 (2006). https://doi.org/10.1109/CVPR.2006.303
11. Kim, W.H., Choi, J.H., Lee, J.S.: Objectivity and subjectivity in aesthetic quality assessment of digital photographs. IEEE Transactions on Affective Computing **11**(3), 493–506 (2020). https://doi.org/10.1109/TAFFC.2018.2809752

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44

12. Kong, S., Shen, X., Lin, Z., Mech, R., Fowlkes, C.: Photo aesthetics ranking network with attributes and content adaptation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision. pp. 662–679 (2016). https://doi.org/10.1007/978-3-319-46448-0_40
13. Leonardi, M., Napoletano, P., Rozza, A.: Modeling image aesthetics through aesthetics-related attributes. London Imaging Meeting **2**, 11–15 (2021). https://doi.org/10.2352/issn.2694-118X.2021.LIM-11
14. Li, L., Duan, J., Yang, Y., Xu, L., Li, Y., Guo, Y.: Psychology inspired model for hierarchical image aesthetic attribute prediction. In: IEEE International Conference on Multimedia and Expo. pp. 1–6 (2022). https://doi.org/10.1109/ICME52920.2022.9859845
15. Li, L., Huang, Y., Wu, J., Yang, Y., Li, Y., Guo, Y., Shi, G.: Theme-aware visual attribute reasoning for image aesthetics assessment. IEEE Transactions on Circuits and Systems for Video Technology **33**(9), 4798–4811 (2023). https://doi.org/10.1109/TCSVT.2023.3249185
16. Li, L., Sheng, X., Chen, P., Wu, J., Dong, W.: Towards explainable image aesthetics assessment with attribute-oriented critiques generation. IEEE Transactions on Circuits and Systems for Video Technology **35**(2), 1464–1477 (2025). https://doi.org/10.1109/TCSVT.2024.3470870
17. Li, X., Li, X., Zhang, G., Zhang, X.: A novel feature fusion method for computing image aesthetic quality. IEEE Access **8**, 63043–63054 (2020). https://doi.org/10.1109/ACCESS.2020.2983725
18. Liu, D., Puri, R., Kamath, N., Bhattacharya, S.: Composition-aware image aesthetics assessment. In: IEEE Winter Conference on Applications of Computer Vision. pp. 3558–3567 (2020). https://doi.org/10.1109/WACV45572.2020.9093412
19. Malu, G., Bapi, R.S., Indurkhya, B.: Learning photography aesthetics with deep cnns. arXiv preprint arXiv:1707.03981 (2017)
20. Nishiyama, M., Okabe, T., Sato, I., Sato, Y.: Aesthetic quality classification of photographs based on color harmony. In: Conference on Computer Vision and Pattern Recognition. pp. 33–40 (2011). https://doi.org/10.1109/CVPR.2011.5995539
21. Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.Y.B., Li, S.W., Misra, I., Rabbat, M.G., Sharma, V., Synnaeve, G., Xu, H., Jégou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P.: Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)
22. Perronnin, F.: Ava: A large-scale database for aesthetic visual analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. p. 2408–2415 (2012)
23. Shu, Y., Li, Q., Liu, S., Xu, G.: Learning with privileged information for photo aesthetic assessment. Neurocomputing **404**, 304–316 (2020). https://doi.org/https://doi.org/10.1016/j.neucom.2020.04.142
24. Soydaner, D., Wagemans, J.: Multi-task convolutional neural network for image aesthetic assessment. IEEE Access **12**, 4716–4729 (2024). https://doi.org/10.1109/ACCESS.2024.3349961
25. Zeng, H., Cao, Z., Zhang, L., Bovik, A.C.: A unified probabilistic formulation of image aesthetic assessment. IEEE Transactions on Image Processing **29**, 1548–1561 (2020). https://doi.org/10.1109/TIP.2019.2941778
26. Zhang, B., Niu, L., Zhang, L.: Image composition assessment with saliency-augmented multi-pattern pooling. In: British Machine Vision Conference. pp. 1–14 (2021), https://api.semanticscholar.org/CorpusID:233168684

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29909-3_44