Supplementary Material
Rethinking Image Aesthetics Assessment: Models, Datasets and Benchmarks

Anonymous IJCAI2022 submission

1 Implementation Details

1.1 Model Structure

The proposed TANet model has three essential components: a *theme understanding network* (TUNet), an *RGB-distribution-aware attention network* (RGBNet) and an *aesthetics perceiving network* (APNet). TUNet is used to perceive image themes and provide guidance on how to extract self-adaptive theme features for prediction. RGBNet can improve the ability to use high-level features extracted from the RGB color space. Finally, APNet can be used to guide the fusion of these features. In addition to two backbones, TANet contains 6 modules, and we describe the layers of each module as follows.

The layers of L^1 are regarded as the adaptive weights of perceiving themes:

- Weights: linear with 365 input size and 100 output size.
- Bias: linear with 365 input size and 1 output size.

The layers of L^2 are used to reduce the spatial redundancy of theme features:

- Linear with 365 input size and 100 output size.
- Batch norm with 100 input size.
- Parametric linear rectified unit (PReLU) with num_parameters = 1 and init = 0.25.
- Dropout with probability = 0.5.

The layers of L^3 are used to fuse the output features (L^2) and weights (L^1) and solve the dynamic dimensions depending on the batch size:

- Linear with dynamic input size and output size.
- Softmax with dimension = 1.
- Unsqueeze with dimension = 2.
- Adaptive average pooling with 10 input size, 4 output size and dynamic stride.
- Squeeze with dimension = 2.

The layers of L^4 can reduce redundancy in the color distribution and similarity features:

- PReLU with num_parameters = 1 and init = 0.25.
- Dropout with probability = 0.75.
- Linear with 20736 input size and 10 output size.

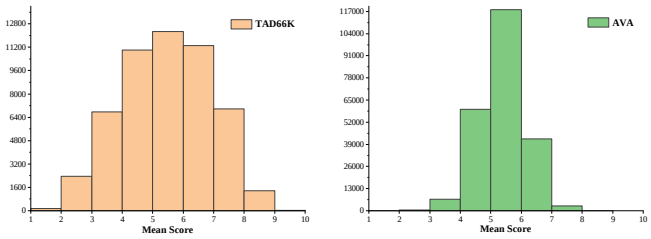

Figure 1: Mean scores of the proposed TAD66K dataset and AVA dataset [Murray *et al.*, 2012]; the distribution of TAD66K is more balanced.

- Softmax with dimension = 1.

We employ L^5 to process the aesthetic features extracted from the backbone (MobileNetV2):

- PReLU with num_parameters = 1 and init = 0.25.
- Dropout with probability = 0.75.
- Linear with 1280 input size and 10 output size.

All features are sent to L^6 to predict an aesthetic score:

- PReLU with num_parameters = 1 and init = 0.25.
- Dropout with probability = 0.75.
- Linear with 30 input size and 1 output size.
- Sigmoid with 10 input size and 10 output size.

1.2 Training Details

On the AVA, FLICKR-AES and TAD66K datasets, the benchmark models in our paper are trained with the recommended parameter settings. To train our TANet, we use Microsoft's neural network intelligence (NNI) tuning tool for training, where the learning rate search space from L^1 to L^6 is set to [0.000001, 0.0000001, 0.0000003], with no decay rate strategy. Specifically, we set the input size to 40 and the number of training epochs to 200, and we take 224 × 224 crops from 256 × 256 fixed images as the input.

2 Dataset Statistics

The proposed TAD66K contains 66,327 images, and the annotation scores range from 1 to 10. We show the distribution of the training set in Fig. 1. We find a serious long-tailed effect in the AVA dataset [Murray *et al.*, 2012], which is mainly

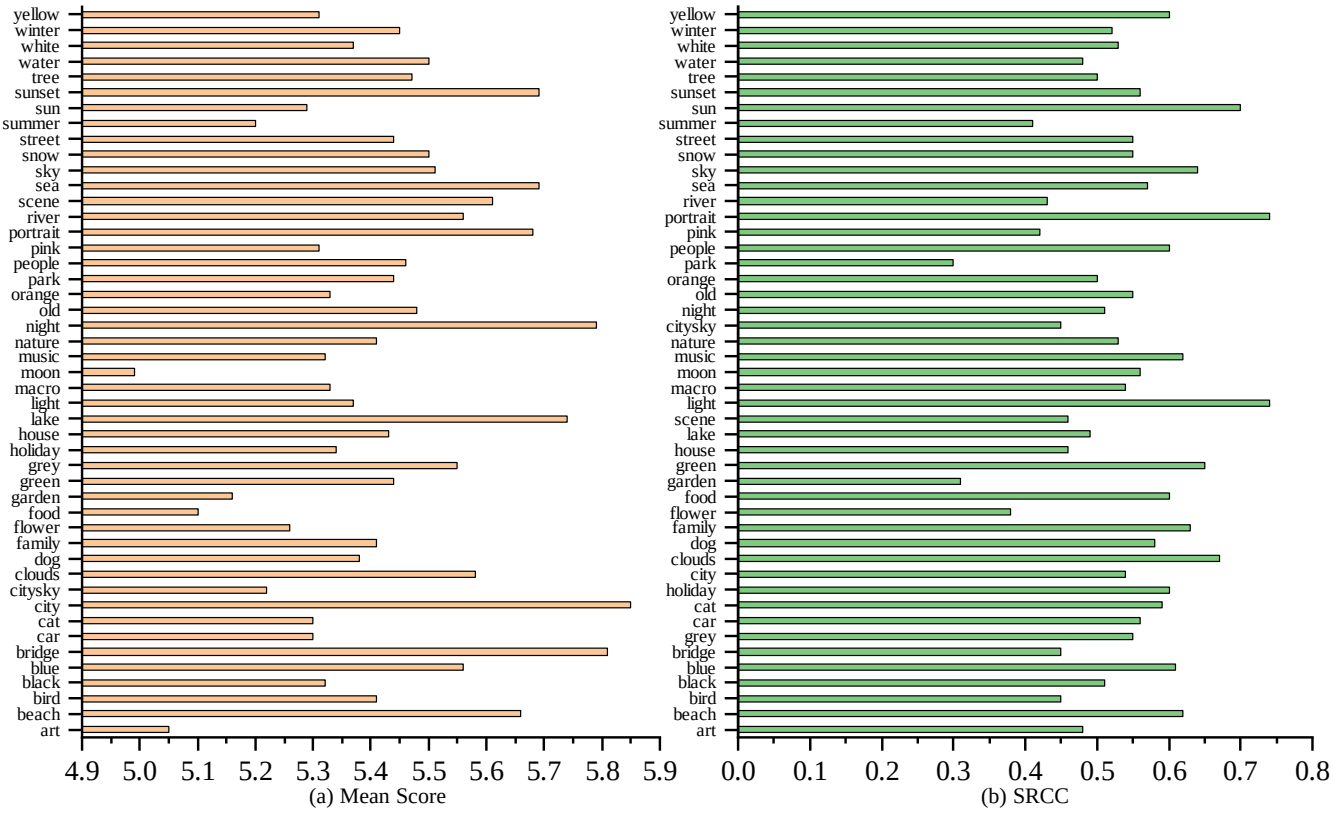

Figure 2: (a) The mean scores of each theme and (b) Spearman's rank correlation coefficient (SRCC) of each theme predicted by TANet.

due to the following two reasons: (1) Multiple themes are mixed together for rating, making it difficult for the annotators to effectively distinguish between images with different themes. Therefore, annotators tend to give an intermediate score. (2) Particularly low-quality images are filtered out at the time of user upload, while particularly high-quality aesthetic images with very strong abstract contexts are difficult to understand by ordinary annotators, and thus, the annotators tend to give an intermediate score if the aesthetic disparity between the images is not obvious. With this in mind, our data collectors tend to make the largest aesthetic distinctions between images at the time of sample collection. Although the annotation scores in the intervals 1∼2 and 9∼10 are still low, we have greatly alleviated the long-tailed effect of the proposed TAD66K.

The mean score of each theme is shown in Fig. 2(a). Some landscape themes, i.e., sunset, lake and sea, are more likely to obtain high scores, whereas some abstract themes, such as art and moon, are more difficult to understand, thus obtaining lower scores. The SRCC predictions for each theme are shown in Fig. 2(b). The model performs poorly for images with mixed themes with objects such as "park" and "garden" and achieves a higher prediction accuracy for themes related to light and shadow such as "sun" and "light".

3 Annotation Details

3.1 Theme Evaluation Criteria

To assist annotators in the annotation process, we provide several criteria for assessing the aesthetics of each theme, as shown in Table 1. We provide these information to help users decide what to focus on when annotating certain themes, but the final annotation result is up to the annotators.

3.2 Annotation Examples

In this section, we show more samples of the TAD66K dataset (Fig. 3). Taking a few typical themes as examples, we can clearly see that images of different themes correspond to different evaluation criteria.

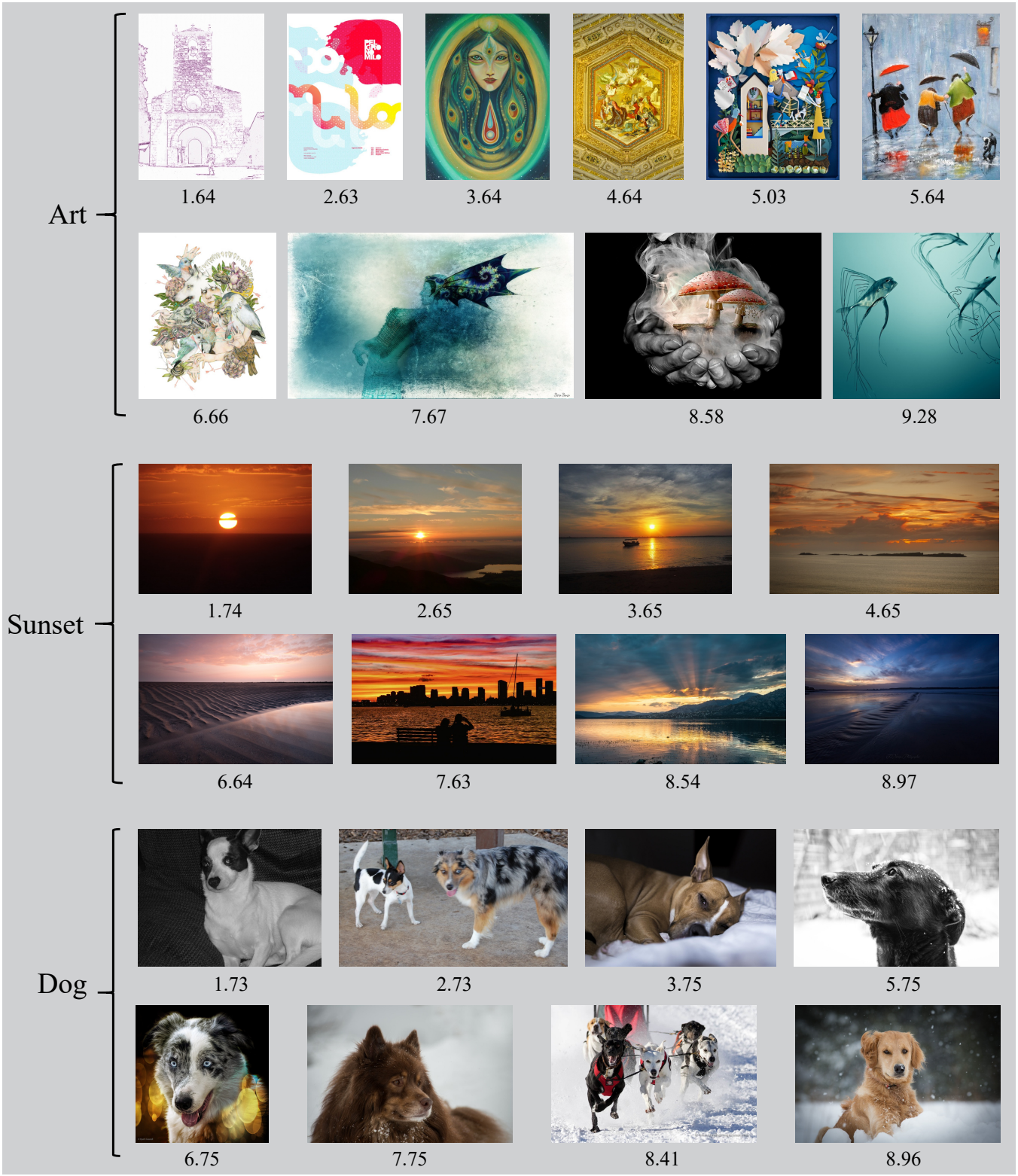

Figure 3: Some samples of the TAD66K dataset with different degrees of aesthetics, people are more likely to give a higher score to well-touched, inspiring or imaginative images.

Table 1: Evaluation criteria for each theme, according to photography, aesthetics, human visual system and other classical theories [Itti and Koch, 2001; Sontag, 2001; Cohen-Or *et al.*, 2006; O'Donovan *et al.*, 2011; Hirsch, 2012; Giblett and Tolonen, 2013; Tormey, 2013; Costello, 2017; Barnbaum, 2017; Kelby, 2019; Kelby, 2020].

Theme	**Evaluation Criteria**
art	Can this image arouse your emotions? Do you feel the image is imaginative? Does the image have profound meaning?
beach	Are the proportions of the elements appropriate? Is the light on the beach comfortable? Do you want to go to the beach in the image for vacation?
bird	Does the bird stand out in the whole image? Does the bird in the image have motion blur?
black	Can the image reflect solemnity? Can the black in the image sets off the content well? Does black affect the sharpness of the image?
blue	Does the color of the image make you feel comfortable? Does the image make you feel quiet? Does the blue in the image have a color cast?
bridge	Is the bridge in the image clear? Does the image reflect the structural characteristics of the bridge? Does the image reflect the majesty of the bridge? Are the bridge and the background properly matched?
gray	Is the main content of the image clear? Would the image be better if it were in color? Can you understand the meaning of this image?
car	Is the car in the image photographed completely or clearly? Does the car in the image show its characteristics?
cat	Is the cat in the image clear? Does the background in the image harmonize with the cat? Is the cat in the image cute? Do you want to have this cat?
holiday	Can you feel the atmosphere of the holiday from the image? Does the image reflect happiness and joy? Do you want to have a holiday like the one in the image?
city	Does the image reflect the characteristics of the city? Can you imagine the life style of the city according to the image? Are you attracted by the city in this image?
clouds	Are the clouds in the image clear? Do the clouds in the image have an artistic sense? Do the clouds in the image make you feel relaxed or warm?
dog	Is the dog in the image clear? Does the background in the image harmonize with the dog? Is the dog in the image cute? Do you want to have this dog?
family	Is each person in the image clearly photographed? Is the proportion of foreground in the image reasonable? Is the view of the camera proper? Does the image make you feel messy? Do you feel the happiness expressed by the image?
flower	Does the color of the flower make you feel comfortable? Are there any local blurring of flowers in the image?

	Are the flowers and the background in harmony? Are the details of the flowers clear enough?
food	Do you like the color of the food? Is the camera angle of the food good? Does the image look messy? Do you think the food in the image is delicious?
garden	Is the overall image harmonious? Does the image look messy? Does the garden in the image make you feel comfortable? Is the light intensity appropriate?
green	Can the image make you feel youthful or natural? Does the green in the image have a color cast? Does the color of the image make you feel comfortable?
house	Is the background of the house clean and harmonious? Does it attract you to want to live? Is the image in a good camera angle?
lake	Does the image reflect the hierarchy and space of the scene? Is the image clear? Can the image make you feel calm and comfortable? Is the proportion of the lake and other scene elements harmonious?
scene	Does the image have a sense of hierarchy? Is the image in a good camera angle?
light	Is the image exposed properly? Does light affect the sharpness of the image? Does the light have the beauty of light and shadow?
macro	Is the main content of the image clear? Is the depth of field of the image appropriate?
moon	Are the details of the moon in the image clear? Is the proportion of the lake and other elements harmonious? Is the moon in the image too dark or too bright?
music	Can this image arouse your emotions? Can you feel the music in the image?
nature	Can the image touch you and make you feel comfortable? Can you feel the beauty of nature in the image? Does the color of the image make you feel comfortable?
citysky	Would you like to live in the city in the image? Can the city in the image reflect a sense of modernity?
night	Is there any noticeable noise in the image? Is the main content of the image clear? Is there any dazzling light in the image? Does the night scene in the image make you feel quiet?
old	Can you feel the history of the image? Does the image have profound meaning?
orange	Can the image make you feel warm or comfortable? Does the orange in the image have a color cast?
park	Does the park in the image appeal to you? Does the image show the characteristics of the park?
people	Are the people in the image clear? Are the people in the right proportion and position in the image?
pink	Is the pink content in the image cute or warm? Does the pink in the image have a color cast?

	Does the color in the whole image make you feel comfortable?
portrait	Does the image show the emotion of the person? Are the people in the image clear? Is the portrait in the right proportion and position in the image? Does the portrait reflect the stereoscopic effect of the face?
river	Does the image reflect the characteristics of the river? Is the composition suitable for the river in this image? Is the proportion of the river and other scenery harmonious?
sea	Can you feel the broadness and grandness of the ocean in the image? Is the proportion of sky, sea and beach appropriate?
sky	Is the sky broad or clean in the image? Is the proportion of the sky and other elements harmonious?
snow	Can you feel the artistic conception of snow in the image? Is brightness appropriate? Is the color monotonous?
street	Does the image make you feel messy? Does the image reflect the characteristics of the street?
summer	Can this image reflect the characteristics of summer? Is the color harmonious? Is the composition of the image appropriate?
sun	Is the shooting time appropriate? Is there any glare in the image? Is the image exposed properly?
sunset	Does the image have overexposed or underexposed areas? Does the image have a peaceful and harmonious beauty? Is the image monotonous? Does the color in the image have a color cast?
tree	Does it show the characteristics of the trees, such as tall, lush and vigorous? Does the image have artistic conception? Does the image have light and shadow features of trees?
water	Does the image reflect the flow of water? Does the image lively and dynamic? Does the image reflect the hierarchy and space of the scene? Is the water in the image clear?
white	Can the image make you feel quiet and pure? Is there any noise in the image? Is the color in the image harmonious?
winter	Does it reflect the characteristics of winter? Is the exposure appropriate? Is the color in the image monotonous?
yellow	Can this image make you feel happy, positive or energetic? Does the color of the image make you feel comfortable? Does the yellow in the image have a color cast?

References

[Barnbaum, 2017] Bruce Barnbaum. *The Art of Photography: A Personal Approach to Artistic Expression*. Rocky Nook, Inc., 2017.

[Cohen-Or *et al.*, 2006] Daniel Cohen-Or, Olga Sorkine, Ran Gal, Tommer Leyvand, and Ying-Qing Xu. Color harmonization. In *SIGGRAPH*, pages 624–630. 2006.

[Costello, 2017] Diarmuid Costello. *On photography: a philosophical inquiry*. Routledge, 2017.

[Giblett and Tolonen, 2013] Rod Giblett and Juha Tolonen. *Photography and Landscape: Photography and Landscape*. Intellect Books, 2013.

[Hirsch, 2012] Robert Hirsch. *Light and lens: Photography in the digital age*. Routledge, 2012.

[Itti and Koch, 2001] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

[Kelby, 2019] Scott Kelby. *The Landscape Photography Book: The step-by-step techniques you need to*. 2019.

[Kelby, 2020] Scott Kelby. *The digital photography book*. Rocky Nook, Inc., 2020.

[Murray *et al.*, 2012] Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *CVPR*. IEEE, 2012.

[O'Donovan *et al.*, 2011] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. Color compatibility from large datasets. In *SIGGRAPH*, pages 1–12. 2011.

[Sontag, 2001] Susan Sontag. *On photography*, volume 48. Macmillan, 2001.

[Tormey, 2013] Jane Tormey. *Cities and photography*. Routledge, 2013.