Learning with Privileged Information for Photo Aesthetic Assessment

Yangyang Shu, Qian Li, Shaowu Liu, Guandong Xu

Advanced Analytic Institute, University of Technology Sydney

Abstract

Privileged information (PI), known as teacher providing students helpful comments, comparisons, and explanations to improve students performance, has been widely applied in various machine learning tasks, resulting in great success. Existing approaches utilizing attributes either fail to leveraging the attributes information thoroughly, or suffer from the complex network structures for automatically attributes learning. Therefore, we propose a new Deep Convolutional Neural Network with Privileged Information (PI-DCNN) for photo aesthetic assessment by utilizing the prior knowledge of photo and photographic elements as privileged information. This paper is the first to systematically summarize all the attributes (i.e., photo and photographic attributes) related to aesthetics assessment. Specifically, we first explore the privileged information of photo and photography attributes, which is available at the training stage but it is not available for the test set. After that, we transfer the probabilistic dependency relations as constraints, and formulate photo aesthetics assessment in a deep convolutional neural network. Lastly, we propose a new pair-wise ranking loss that can exploit the relationship of photo aesthetic quality within a pair of photos. Experimental results on two widely benchmark databases of aesthetic assessment, `AADB` and `AVA`, demonstrate the effectiveness of the proposed PI-DCNN method on photo aesthetic assessment task.

Keywords: photo aesthetic assessment, privileged information, ranking algorithm

1. Introduction

Photo aesthetic assessment with a wide range of applications in photographic art, has attracted more and more attention in recent years due to the increasing

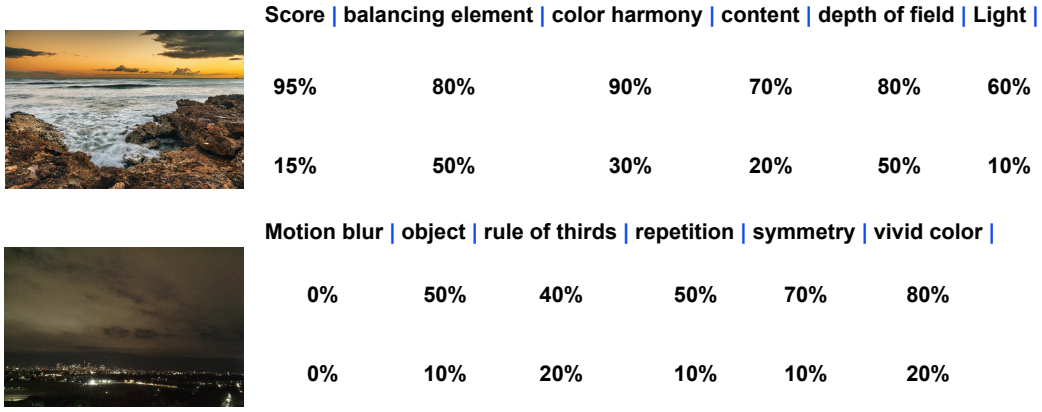


Figure 1: Photos from AADB database [1] obtain different aesthetic scores due to different attributes. Note that aesthetic scores and attributes are written as the percentages for simplicity. There are six attributes (the first row) having positive correlation with the aesthetic score, i.e., high values of attributes generate high values aesthetic scores. This observation motivates us to exploit the dependency relationship between photo aesthetic assessment and aesthetic attributes.

requirements for art appreciation [2][3]. Aesthetic assessment is a subjective task that heavily relies on human perception to the photos. From a computational perspective, mathematically quantify perception is very difficult and thus poses a great challenge for aesthetics assessment. Consequently, the research of computational aesthetics may still be at the early stage of development [2, 4]. Researchers have found that human perception can be commonly affected by different usages of psychological rules and photographic factors including lighting [5], contrast [6], composition [7] and photo content [8] etc. Figure 1 gives an example of high-quality photo and low-quality photo with their some high-level photographic attributes respectively. In this example, the photo with high aesthetic has nice attributes with good lighting and vivid color, which makes it fascinating, while the low aesthetic photo has some attributes with poor lighting and dull color. Based on above observations, a more realistic approach of aesthetics assessment is to exploit the influences of such photographic factors so as to predict aesthetics results.

Existing work introducing describable photo attributes into aesthetics assessment is popular and common, for example multi-task framework [9] addresses the correlation issue between automatic aesthetic quality assessment and semantic recognition. The semantic information is considered key to help discover representations for aesthetic quality assessment. Similar works utilizing single at-

tribute are common, such as semantic-aware hybrid Network (SANE) [10], color harmony-based aesthetic quality network [11] and composition-aware image aesthetic network [12]. All of these approaches focus on only one attribute i.e. semantic information, color harmony or composition et. assisting to address assessment problem. Meanwhile, multi-attribute framework in photo aesthetic assessment has been proposed, for example, eight-attributes deep convolution neural network (DCNN) [13] is proposed to learn the aesthetic score and attributes jointly by using a deep convolution network with a merge-layer. More multi-attribute networks have been proposed in current work, such as brain-inspired deep networks (BDN) [14], rating pictorial aesthetics using deep learning system (PAPID) [15], user-friendly aesthetic ranking framework (USAR) [16] and multi-level spatially pooled (MLSP) features architectures [17] etc.

All of above work either resort to hand-crafted feature extraction or automatic feature extraction by deep learning, which have mainly three limitations in common. First, methods based on hand-crafted features are ineffective, due to the exhaustive repeats of hand-crafted feature extraction for each aesthetic assessment task. Second, deep networks based methods avoid the hand-crafted feature extraction but suffer from complex network structures and high computation cost. This mainly because an extra deep network is required to automatically learn high-level features before the aesthetic assessment[14][13]. Third, both all methods of above two categories merely focus on some photo attributes even one photo attribute. However, photographic factors including color and location of object categorization from content are also vital, but are usually ignored[9][10][1][11][12]. Namely, inherent dependencies or correlations between photo attributes and photographic attributes are not fully exploited by previous methods.

It is well-known that professional photographers use different photographic techniques and have different aesthetic criteria in mind when taking different types of photos [18]. Namely, this additional information (aesthetic criteria and photographic techniques) is informative for aesthetic assessment than the traditional training data alone. However, this additional information can not be fully exploited by traditional machine learning method, because some of them can be only observed by photographers, such as the HDR and macro captured by the professional cameras. The new supervised learning paradigm, namely learning using privileged information (LUPI), can be used to solve this problem. Inspired by this, our paper utilizes the privileged information for accelerating the feature learning process and thus improving aesthetic score prediction. We exploit 18 different types of privileged information in the context of aesthetic assessment, including balancing elements, color harmony, complementary, vivid color, content, depth

of field, light, motion blur, object, rule of third, duotones, HDR, long exposure, macro, negative photo, silhouettes, soft focus and vanishing point.

Therefore, in this paper, we propose a novel attributes-aware photo aesthetic assessment method, where attributes as privileged information are used in training but not available in testing [19]. Specifically, first, we systemically summarize photo-based and photography-based attributes from aesthetics and photographic research. Second, we successfully infer probabilistic dependencies between attributes and aesthetics from the summarized prior knowledge, and further transfer them as constraints of aesthetic recognition and regression. Third, we adopt pairwise photos to explicitly exploit the relation of photo pairs based on our ranking algorithm. Fourth, we conduct a number of experiments on two widely used photo aesthetic assessment benchmarks, AADB [1] and AVA [20], and the experimental results demonstrate significant improvements over existing state-of-the-art methods. Our contributions can be summarized as follows.

- As a comprehensive study to explore the aesthetics from photo and photography, this paper summarizes the relationship between aesthetic score and aesthetic attributes.
- This paper takes main aesthetic attributes as privileged information to demonstrate the feasibility of the proposed rating photos method enhanced via privileged information. Specifically, we first infer probabilistic dependencies between main aesthetic attributes and aesthetics from the summarized art and photography theory. Then in our PI-DCNN model, we transfer the privileged information to constraint and formulate photo aesthetic assessment as a constrained optimization problem.
- We improve a ranking algorithm to utilize different pairwise photos for training model. We show this algorithm substantially improves the performance via setting loss function in the fully connected layer of the network.

2. Related Works

2.1. Photo Aesthetic assessment Under attributes

A comprehensive survey related to recent computer vision techniques used in the assessment of photo aesthetic quality can be found in [4, 2]. In this section, we focus on several works that utilize attributes for photo aesthetic assessment.

In the early work, people try to find and design some features which are assumed to model the photographic/artistic aspect of photos in order to distinguish

photos of different qualities. Datta *et al.*[18] extracted certain visual features based on the intuition including a low depth-of-field indicator, a colorfulness measure, a shape convexity score and a familiarity measure that they can discriminate between aesthetically pleasing and displeasing photos. Ke *et al.*[21] proposed three distinguishing factors, simplicity, realism, and basic photographic technique, making a photo high-quality or low-quality. Then, spatial distribution of edges, color distribution, hue count, blur, contrast and brightness are designed. More similar works can be found in[22, 23]. Although their manually designed features are relevant to photography techniques, the high-level semantic attributes are not fully captured.

In recent years, some work just used one or few attributes in their models due to lacking of summary and theory of aesthetic attributes. For example, Kao *et al.*[9] proposed a multi-task deep learning framework to addresses the correlation issue between automatic aesthetic quality assessment and semantic recognition. They argued that semantic recognition task offers the key to address automatic aesthetic quality assessment. Cui *et al.*[10] designed a novel semantic-aware hybrid network (SANE), which captures the information from object categorization and scene recognition to improve the accuracy of photo aesthetics assessment. Zhang *et al.*[24] proposed a Gated Peripheral-Foveal Convolutional Neural Network (GPF-CNN) to exploit the semantic information for photo aesthetic assessment. Kong *et al.*[1] proposed to learn a deep convolutional neural network to rank photo aesthetics in which photo content information is utilized in photo aesthetics rating problem. Above-mentioned three works focus on analyzing the semantic and content information of a photo such as semantic labels: "Sky" and "Architecture". However, its color and position also provide important information but have been ignored. Nishiyama *et al.*[11] proposed a "bags-of-color-patterns" method for aesthetic quality classification with the help of the color harmony of photos. Mai *et al.*[25] proposed composition-preserving deep network and Liu *et al.*[12] proposed to model the photo composition information as the mutual dependency of its local regions, and design a novel architecture to leverage such information to boost the performance of aesthetics assessment. However, their method only used one attribute, color harmony or composition, ignoring other important attributes such as content, which is crucial for photo aesthetic assessment.

To exploit attributes thoroughly, some other works utilize multiple attributes for aesthetic assessment. In their works, the attributes are used in not only training, but also testing for measuring aesthetic score. For example, Lu *et al.*[15] adopt attributes to allow unified feature learning and classifier training for photo aesthetic assessment. Lv *et al.*[16] proposed to generate an aesthetic distribution

that integrate dozens of aesthetic attributes for all input photos . Then, by concatenating distribution of each photo, a final aesthetic distribution of the user is released. For their method, attributes as input are needed in training. During testing, the attributes are still typically firstly predicted by model, or captured by database. This is usually complex and hardly satisfied in reality.

Some recent work avoid use attributes in testing. They designed additional deep network structure to automatically generate high-level aesthetic attributes. For example, Wang *et al,*[14] designed Brain-Inspired Deep Networks (BDN) to first learns attributes through the parallel supervised pathways. Then, they associated and transformed those attributes into the overall aesthetics rating for this task. Malu *et al.*[13] proposed a novel multitask deep convolution neural network (DCNN), which jointly learns eight aesthetic attributes along with the overall aesthetic score. They used a deep convolution network with a merge-layer. The merge-layer collects pooled features of the convolution maps, and the aesthetic score and attributes are learned based on the merge-layer. However, in these methods, they designed some branches of frameworks to learn attributes, which results in high complexity of the network.

2.2. Learning Under Privileged Information

Vapnik and Vashist [19] firstly proposed a new learning paradigm i.e., using privileged information (LUPI) where at the training stage a teacher gives some additional information, while these information is not available at the test stage. Then privileged information has been applied to various computer vision task. You *et al.*[26] apply depth features from depth photos that are captured by depth cameras as privileged information to improve face verification and person re-identification in the RGB photos. Sarafianos *et al.* [27] utilized privileged information in a regression-based method to estimate the height using human metrology. Lambert *et al.*[28] propose a new Learning Under Privileged Information algorithm to use a heteroscedastic dropout and make the variance of the dropout a function of privileged information.

Unlike previous work, we first summarize main aesthetic attributes from photo and photography according to aesthetics and photography theory. Then, we regard these aesthetic attributes as privileged information, which are not necessary in testing. Instead of designing additional network to predict the values of attributes, we leverage the relationship between aesthetic score and aesthetic attributes and infer this probabilistic dependencies as a constraints applied in loss function. In summary, we introduce an analytical method to systematically incorporate the privileged information for photo aesthetic assessment.

Table 1: The dependencies between 18 aesthetic attributes and the aesthetic quality of photos. Positive represents higher value of the attribute and negative represents lower value of the attribute. $\surd$ indicates the existence of high dependency of the aesthetic score on the aesthetic attribute.

photo-based attributes		high quality	low quality	**photography-based attributes**		high quality	low quality
balancing elements	positive	$\surd$		duotones	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
color harmony	positive	$\surd$		HDR	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
complementary	positive	$\surd$		long exposure	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
vivid color	positive	$\surd$		macro	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
content	positive	$\surd$		negative photo	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
depth of field	positive	$\surd$		silhouettes	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
light	positive	$\surd$		soft focus	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
motion blur	positive	$\surd$		vanishing point	positive	$\surd$	
	negative		$\surd$		negative		$\surd$
object	positive	$\surd$					
	negative		$\surd$				
rule of third	positive	$\surd$					
	negative		$\surd$				

3. Attributes as Privileged Information in Photo Aesthetics

The aesthetic attributes of photos are usually exploited from the perspective of photo and photography. Attributes from photo mean the robust feature representations describing the aesthetic aspect of a photo, such as color [21], content [29], light [30], depth of field [29], rule of third [29] and motion blurs [31] etc. Such attributes are related to the artistic aspect of photos and can evaluate the quality of photos. Attributes from techniques for photography focus on the photographic aspects of photos in order to distinguish photos of different aesthetics, such as duotones [32], long exposure [21], silhouettes [21] and soft focus [33] etc. These attributes usually get involved with photography rules.

Table 1 summarizes the photo-based and photography-based attributes, where the dependencies among these attributes are also presented.

3.1. Photo-based Attributes

Ten attributes are used to exploit the relationship between attributes and aesthetic quality from photo perspective in our work. They are balancing elements, color (color harmony, vivid color and complementary), content, depth of field, light, motion blur, object and rule of thirds respectively shown in Figure 2.

Balancing element is vital for the aesthetic quality of photos, which is a fundamental principle of visual perception in that the eye seeks to balance the elements

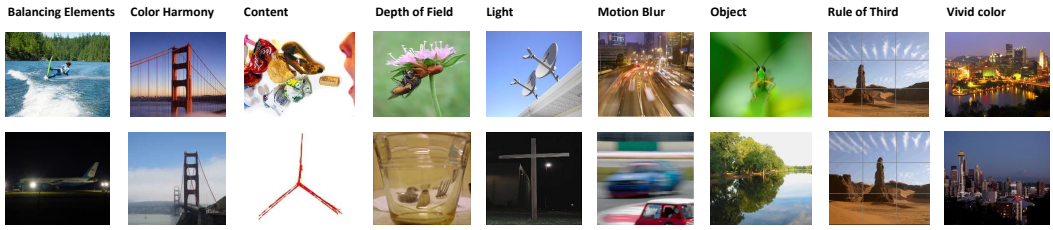


Figure 2: Different aesthetic qualities w.r.t. different photo attributes as privileged information. Top row includes photos with high aesthetic qualities. Bottom row includes photos with low aesthetic qualities.

and establish the harmony in a photograph[34]. Photos composition organizes the positions of objects within the photo and balances them w.r.t. lines or points so as to achieve the harmony. Therefore, by taking the design and rule of photos into account, the aesthetic scores of photos are more likely to be high when having a good balancing element, and tend to become low while without a good balancing element.

Colors involves *color harmony*, *vivid color* and *complementary color*, which has a strong relationship with aesthetic quality. This attribute can identify the differences in the color palette used by professional photographers and non photographers [21]. Color harmony is the term for colors that are thought to match. In other words, colors that look aesthetically pleasing side-by-side [35]. Vivid color refer to an intense feeling, or a photo in your mind that is so clear you can almost touch it. Complementary colors are pairs of colors which, when combined or mixed, cancel each other out (lose hue) by producing a grayscale color like white or black [36]. The photos taken by professional photographers are more colorful than the ones by non-professional photographers. Thus, vivid color, good color harmony and appropriate complementary color are more likely to produce photos with the high aesthetic quality.

Object and *rule of third* are another two important features among photo attributes or photographic attributes. In Dhar *et al.*[29]'s research, they predict whether a photo contains some large objects well separated from its background. They also found some photos with one or more salient are usually high-quality. Rule of third means a photo can be marked two equally spaced horizontal lines and two equally spaced vertical lines. This rule divided photo into nine equal parts and some important compositions and objects should be placed along these lines or intersections. In Dhar *et al.*[29]'s research, they utilize the salient object detector to calculate the rule of third and found that it will be more aesthetically

Figure 3: Different aesthetic qualities w.r.t. different photo-based attributes. Top row includes photos with high aesthetic qualities. Bottom row includes photos with low aesthetic qualities.

pleasing to place the main subject of the picture on one of two vertical and two horizontal lines or on one of their intersections. Generally speaking, presence of a salient object and photo with rule of third can induce high quality of aesthetic photo.

Content attributes refer to the presence of specific objects including human faces, animals, and scene. Dhar *,et al.*[29] found that the aesthetic score given by human is very susceptible to being influenced by the content in the photo. Therefore, diverse content is strongly correlated to high aesthetic score of photo.

The attribute of *low depth of field (DoF)* is one photo where objects within a small range of depths in the world are captured in sharp focus, while objects at other depths are blurred (often used to emphasize an object of interest). Dhar *et al.* [29] concluded that shallow or low depth of field is more likely lead to high aesthetic score of photo, while deep or high depth of field is more likely lead to low aesthetic score of photo.

Light attribute in terms of intensity is important factors for aesthetic quality evaluation. Light intensity refers to the strength or amount of light produced by a specific lamp source. As a general rule, if the lightness difference between the brightest and darkest regions of a photograph is small, the photo can be perceived as under- or over-saturated and washed-out photo, which makes the aesthetic quality of photos worse [30]. In general the high difference in lightness in photos can contribute to high aesthetic score of photo.

Motion blur is the apparent streaking of moving objects in a photograph or a sequence of frames, such as a film. It occurs when the photo is recorded changes during the recording due to rapid movement or long exposure. Thus, under moving objects in a photograph condition, a photo is more likely to invoke high aesthetic quality of photo when it contains motion blur [31].

3.2. Photography-based Attributes

Complementary to the photo attributes, we introduce eight photography-based attributes shown in Figure 3. They are duotones, high dynamic range(HDR), long exposure, macro, negative photo, silhouettes, soft focus, vanishing point respectively.

Duotone refers to a photo with various shades of a hue mapped in a vector through a color space. The colorant, the gradient curve, and the number of colorants are used to define the slice through the color space. Photos are printed with two or more analogue colorants[32]. Photos with duotone are more likely to produce the photos with high aesthetic quality.

High dynamic range (HDR) used to reproduce a greater dynamic range of luminosity than what is possible with standard photographic techniques. This attribute is often used for display devices, photography, 3D rendering, and sound recording including digital imaging and digital audio product[37]. In Reinhard *et al.*[38]'s research, they found the compositing and tone-mapping of photos with HDR more probably lead to the high aesthetic quality of photos.

Long-exposure involves using a long-duration shutter speed to sharply capture the stationary elements of photos while blurring, smearing or obscuring the moving elements. Long-exposure photography captures one element that can not captured by conventional photography. Ke *et at.*[21] found that a photographer uses a long shutter speed to capture a motorial object, which is more likely to invoke the high aesthetic quality of photos.

Other cameras techniques like *macro* photography and *soft focus* have an extensive application in photo shoot. Macro photography extreme is used in close-up photography, usually of very small subjects and living organisms like insects. The size of the subject in the photograph is greater than life size. In photography, soft focus is a lens flaw, in which the lens forms photos that are blurred due to spherical aberration. A soft focus lens deliberately introduces spherical aberration in order to give the appearance of blurring the photo while retaining sharp edges; it is not the same as an out-of-focus photo, and the effect cannot be achieved simply by defocusing a sharp lens. Soft focus is also the name of the style of photograph produced by such lens. Hence, generally speaking, photo photographed with photography and soft focus skills has a higher aesthetic quality[39][33].

In photography, a *negative photo* [40] is additionally color-reversed, with red areas appearing cyan, greens appearing magenta and blues appearing yellow, and vice versa. *Silhouettes* [21] is the photo of a person, animal, object or scene represented as a solid shape of a single color, usually black, with its edges matching

the outline of the subject. Finally, *vanishing point* [41] refer to a point on the photo plane of a perspective drawing where the two-dimensional perspective projections (or drawings) of mutually parallel lines in three-dimensional space appear to converge such as the scene with railway or road. These three attributes, i.e. negative photo, photo with silhouettes and vanishing point usually can result in high aesthetic quality of photos [40][21][41].

4. Preliminary

Before proposing our method, this section first formalizes the problem of learning regression model from the labeled data along with privileged information. The goal of regression learning is to learn function $f : \mathbf{X} \rightarrow \mathbf{Y}$, which mapping from feature space $\mathcal{X}$ to a label space $\mathcal{Y}$. Denote a set of triples $S = \{\mathbf{X}, \tilde{\mathbf{X}}, \mathbf{Y}\}$, where $\mathbf{X} \in \mathbb{R}^{N*d}$ is feature matrix and $\tilde{\mathbf{X}} \in \mathbb{R}^{N*\tilde{d}}$ is privileged information matrix. N is the number of samples. d and $\tilde{d}$ are the dimensions of feature and privileged information respectively. $Y \in \mathbb{R}^N$ is the ground-truth vector, which is a aesthetic score given by human annotation from databases.

During training phase, we sample the data $\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}$ from S following the distribution $(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y}) \sim p(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y})$. Nevertheless, sampling data from test dataset that follows $\mathbf{x} \sim p(\mathbf{x})$ with unknown aesthetic score $\mathbf{y}$ and privileged information $\tilde{\mathbf{x}}$. Our aim is to solve the following optimization problem:

$$\min_{\theta} \mathbb{E}_{\mathbf{x_i}, \tilde{\mathbf{x}}_\mathbf{i}, \mathbf{y_i} \sim \mathrm{p}(\mathbf{x}, \tilde{\mathbf{x}}, \mathbf{y})} [\ell(\mathbf{y}, g(\mathbf{x}, \tilde{\mathbf{x}}, \theta))] \tag{1}$$

where θ is the parameter of the function g is our network mapping from photos to aesthetic scores, $\mathbf{y}$ is the ground truth, $\ell(\cdot, \cdot)$ is a loss function.

The framework of the proposed method is shown in Fig. 4. During training, we first learn the deep network model by using photos from training sets. Since our model predicts a continuous aesthetic score other than category labels, we replace the softmax loss with regression loss function and set one node in the last layer as aesthetic score. Inspired by [42], we start by fine-tuning the deep residual network (ResNet) [43] using different losses to predict aesthetic scores. The aesthetic attributes (Att_fea in Fig. 4) are used as privileged information to help model constructing better feature representations for aesthetic assessment. Then during test, we use the trained deep network model to predict the aesthetic score of an unknown photo.

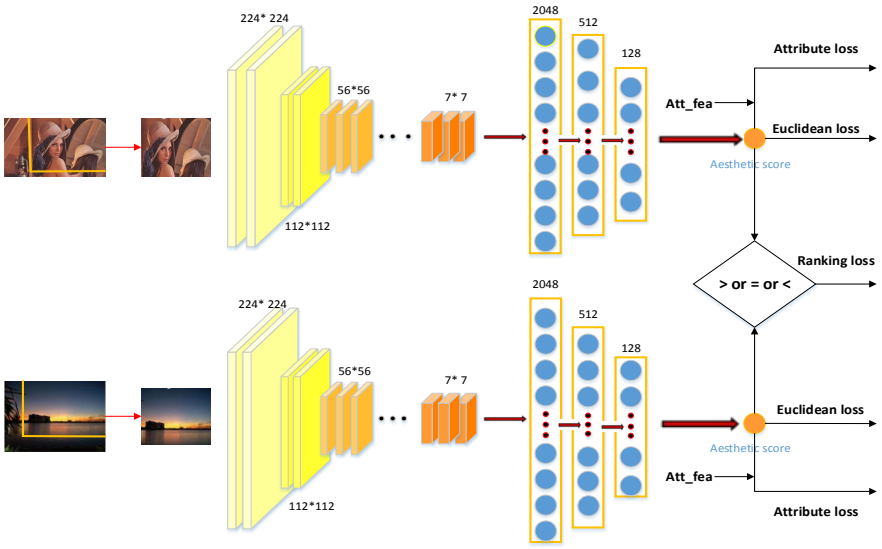


Figure 4: An overview of our proposed deep learning framework for photo aesthetic assessment. "Att_fea" are aesthetic attributes used as privileged information. Our model utilizes the ResNet architecture is augmented by replacing the top softmax layer with a Regression loss. We also adopt ranking loss additionally for model training. To produce a privileged information-aware loss for aesthetic assessment, our model also considers different importance of attributes by combining different weights.

5. PI-DCNN for Photo Aesthetic Assessment

Denote three tuples $S = \{(\mathbf{x_i}, \tilde{\mathbf{x}}_\mathbf{i}, y_i)|i = 1, ..., N\}$, where $\mathbf{x_i} \in \mathbb{R}^d$ represents a color photo from training database. $\tilde{\mathbf{x}}_\mathbf{i} \in \mathbb{R}^K$ denotes a photo with K-dimensional aesthetic attributes. Each aesthetic attribute is binary $\tilde{x_i}^k \in \{0, 1\}$ and $y_i \in \mathbb{R}$ represents aesthetic scores and N is the number of training sets. The goal is to learn a deep convolutional neural network $\hat{\mathbf{y}} = G(\mathbf{x}, \theta_G)$, where $\hat{\mathbf{y}} \in \mathbb{R}^N$ represents the predicted label. θ_G denotes the parameters in this deep network. The object function of this deep network is defined as follow:

$$\min_{\theta_G} \sum_{i=1}^{N} L_{reg}(\hat{y}_i, y_i) + \sum_{i=1}^{N}\sum_{j=1}^{N} C_1 L_{rank}(\hat{y}_i, \hat{y}_j, y_i, y_j) + \sum_{i=1}^{N}\sum_{k=1}^{K} C_2 L_{pi}(x_i, \tilde{x_i}^k, \hat{y}_i), \tag{2}$$

where C_1 and C_2 are two coefficients, $L_{reg}(\hat{y}_i, y_i)$ denotes the basic loss function, $L_{rank}(\hat{y}_i, \hat{y}_j, y_i, y_j)$ represents the ranking loss function and $L_{pi}(x_i, \tilde{x_i}^k, \hat{y}_i)$ is the

loss function of aesthetic attributes as privileged information for rating photos.

Three loss functions perform optimization from different perspective. Regression loss makes the continuous aesthetic scores of our model closing to real scores. Ranking loss is used to explicitly exploit relative rankings and difference of photo pairs. Combination regression loss and ranking loss, not only do two predicted photos scores approach to their real scores respectively, but also the ranking and gap between them are closing to accuracy. Privileged information loss utilize some domain knowledge and additional information between attributes and aesthetics, which can be exploit in our proposed model. Hence, we combine the loss functions in our proposed model.

5.1. Privileged Information Loss

Photo attributes and photographic attributes are highly related to photo aesthetic assessment. If a photo has these obvious attributes than others, the photo is more possible to have a higher aesthetics[29]. Therefore, we can infer the probabilistic dependencies between attributes and aesthetics as:

$$\begin{aligned} p(\hat{y} \geq C|\tilde{x_i}^k = 1) > p(\hat{y} < C|\tilde{x_i}^k = 1) \\ p(\hat{y} < C|\tilde{x_i}^k = 0) > p(\hat{y} \geq C|\tilde{x_i}^k = 0) \end{aligned} \tag{3}$$

where $p(\hat{y} \geq C|\tilde{x_i}^k = 1)$ and $p(\hat{y} < C|\tilde{x_i}^k = 1)$ indicate the probabilities of high aesthetic score $\hat{y} \geq C$ and low aesthetic score $\hat{y} < C$ respectively, when observing the obvious attributes. C is the threshold for dividing the photos into high quality and low quality. $\tilde{x_i}^k = 1$ represents the consistency in this attribute. $p(\hat{y} < C|\tilde{x_i}^k = 0)$ and $p(\hat{y} \geq C|\tilde{x_i}^k = 0)$ represent the probabilities of low aesthetic score and high aesthetic score respectively, when observing the low value of the attributes. $\tilde{x_i}^k = 0$ represents the violation of this attributes. Specifically, if some attributes are unknown or missing, their probability expressions are discarded.

Our model adopts ReLU function to penalize the samples violating the probabilistic dependency relationship. The corresponding penalty $l_i(x_i, \tilde{x_{ii}^k}, \hat{y_i})$ is encoded from privileged information according to Eq. 3 as below:

$$\begin{aligned} &\ell_i(x_i, \tilde{x_i}^k, \hat{y_i}) = \tilde{x_i}^k * [p(\hat{y_i} < C|\tilde{x_i}^k = 1) - p(\hat{y_i} \geq C|\tilde{x_i}^k = 1)]_+ \\ &+ (1 - \tilde{x_i}^k) * [p(\hat{y_i} \geq C|\tilde{x_i}^k = 0) - p(\hat{y_i} < C|\tilde{x_i}^k = 0)]_+ \\ &= \tilde{x_i}^k * [1 - 2 * p(\hat{y_i} \geq C|\tilde{x_i}^k = 1)]_+ + (1 - \tilde{x_i}^k) * [2 * p(\hat{y_i} \geq C|\tilde{x_i}^k = 0) - 1]_+ \end{aligned} \tag{4}$$

where $[\cdot]_+ = max(\cdot, 0)$ is ReLU function. According to the property of logistic regression, we apply sigmoid function to replace the probabilistic dependencies

between attributes and aesthetics label as in Eq. (5)

$$\begin{aligned} p(\hat{y} \geq C|\tilde{x}^k) &= \sigma(\hat{y} - C) \\ p(\hat{y} < C|\tilde{x}^k) &= 1 - \sigma(\hat{y} - C) \end{aligned} \tag{5}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$. Thus, we rewrite Eq.4 as

$$\begin{aligned} &\ell_i(x_i, \tilde{x_i}^k, \hat{y_i}) = \tilde{x_i}^k * [1 - 2 * p(\hat{y_i} \geq C|\tilde{x_i}^k = 1)]_+ \\ &+ (1 - \tilde{x_i}^k) * [2 * p(\hat{y_i} \geq C|\tilde{x_i}^k = 0) - 1]_+ \\ &= \tilde{x_i}^k * [1 - 2 * \sigma(\hat{y_i} - C)]_+ + (1 - \tilde{x_i}^k) * [2 * \sigma(\hat{y_i} - C) - 1]_+ \end{aligned} \tag{6}$$

Since different attributes as privileged information have different influence on aesthetic rating, we set each privileged information with corresponding weight coefficient. Finally, we define $loss_{pi}$ as

$$loss_{pi} = \frac{1}{2N} \sum_{i=1}^{N} \sum_{k=1}^{K} \alpha^k \ell_i(x_i, \tilde{x_i}^k, \hat{y_i}) \tag{7}$$

where α^k is the weight coefficient between different attributes and aesthetic scores. This denote the different importance between attributes and aesthetic score. The details are showed in Fig. 7.

5.2. Ranking Loss

Differ from privileged information loss, ranking loss describes relative rankings of photo pairs in which the ranking of photo aesthetics are directly modeled in the loss function. Specifically, if the aesthetic score of photo-i is higher than the aesthetic score of photo-j, the corresponding estimated score of photo-i should be still higher than photo-j. This is given in Eq. 8 as:

$$loss_{rank} = \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{N} \{(y_i - y_j) - (\hat{y_i} - \hat{y_j})\}^2 \tag{8}$$

where y_i and y_j are the real label of photo-i and photo-j respectively. $\hat{y_i}$ and $\hat{y_j}$ are the corresponding predicted label in deep network model. The equation makes the readily available ordinal information and satisfies the assumption stated in the beginning of Section 5.2.

In Kong*et al.* [1]'s work, they also proposed their ranking algorithm. We compare our ranking algorithm with their. The ranking algorithm in Kong *et al.*'s method is shown in Eq. 9.

$$loss_{rank} = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N} max\{0, \beta - \delta(y_i \geqslant y_j)(\hat{y_i} - \hat{y_j})\} \tag{9}$$

where $\delta(y_i \geqslant y_j) = \begin{cases} 1, & if \quad y_i \geqslant y_j \\ -1, & if \quad y_i < y_j \end{cases}$, and β is a specified margin parameter. Compared with Kong's ranking method, we have the following improvements.

- Firstly, Kong's method uses piecewise function to achieve ranking algorithm, which suffers from the high model complexity. Instead of binary function $\delta(y_i \geqslant y_j)$ in Kong's ranking method, we adopt difference function $(y_i - y_j) - (\hat{y_i} - \hat{y_j})$ to finish ranking algorithm. This transfers the sign function into the continuous function, which improve the efficiency in deep network.

- Secondly, the ranking model in Kong's method merely considers the ranking relation among different photos, while our method additionally considers the gap (difference) between the score of predicted photo pairs and real photo pairs, i.e., $(y_i - y_j) - (\hat{y_i} - \hat{y_j})$. For example, we have photo-i and photo-j. The aesthetic score of photo-i is higher than the aesthetic score of photo-j. Kong's ranking algorithm considers the estimated score of photo-i should be still higher than photo-j but fails to measure such difference. This is important for improving the prediction of our proposed network.

5.3. Regression Loss

To make the predicted label $\hat{y}$ be approximate to the real label y, we define the following formula:

$$loss_{reg} = \frac{1}{2N}\sum_{i=1}^{N} ||\hat{y_i} - y_i||_2^2 \tag{10}$$

where y_i is the average ground-truth rating for photo-i, and $\hat{y_i}$ is the estimated score by the CNN model.

Hence, we combine Regression loss denoted by $loss_{reg}$, ranking loss denoted by $loss_{rank}$, privileged information loss denoted by $loss_{pi}$ as rating loss:

$$loss = loss_{reg} + C_1 loss_{rank} + C_2 loss_{pi} \tag{11}$$

Algorithm 1 gives the learning procedure of our proposed method.

Algorithm 1 The learning algorithm of proposed PI-DCNN framework

Input:

Training sample $(\mathbf{x_i}, \tilde{\mathbf{x}}_\mathbf{i}, y_i)$, $(\mathbf{x_j}, \tilde{\mathbf{x}}_\mathbf{j}, y_j)$, i=1,...,$N$

Coefficient α, α^k, learning rate η, bitch size m, threshold value C

Randomly initialize parameters of Model parameters θ_G;

repeat

for each training sample $(\mathbf{x_i}, \tilde{\mathbf{x}}_\mathbf{j}, y_i)$ in a mini-batch of m training photos **do**

Calculate the Regression loss as Eq.10

Calculate the ranking loss as Eq.8

Calculate probabilistic dependencies $p(\hat{y} \geq C|\tilde{x}^k), p(\hat{y} < C|\tilde{x}^k)$ as Eq.5

end for

Update the deep network by gradient descent $\theta_G := \theta_G - \eta \frac{\partial G_G(\theta_G)}{\partial \theta_G}$

until Converges

Output:

Aesthetic scoes $\{y_i = \theta_G^T x_i | i = 1, ..., N\}$

6. Experiment

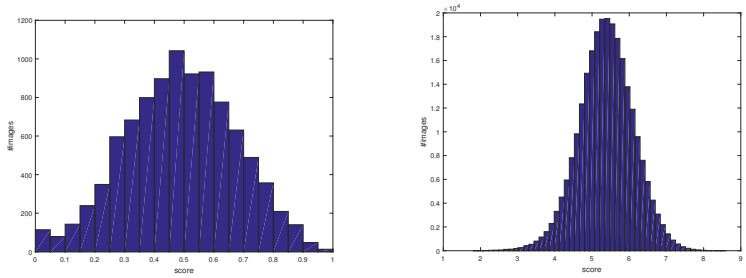
To evaluate the effectiveness of our proposed PI-DCNN model, We conduct experiments on two benchmark databases: Aesthetics and Attributes database (AADB) [1] and Aesthetics Visual Analysis database (AVA) [20].

6.1. Databases

AADB database contains a set of 10,000 photographic photos downloaded from the Flickr website [1]. Aesthetic quality score and eleven attributes are provided by five different individual raters using Amazon Mechanical Turk 11 attributes in this database are balancing elements, color harmony, content, depth of field, light, motion blur, object, repetition, rule of third, symmetry and vivid color respectively. The aesthetic quality scores of photos in this database are showed in Fig.5(a) and the details of the eleven numerical aesthetic attributes are summarized in Table 2. We find that the distributions of the aesthetic scores are approximately Gaussian. Then we count and list the distribution of eleven attributes' values in Fig.6. As can be seen from Fig.6, we divide the attributes of photo into three groups, including positive (above threshold), null (equal threshold) and negative (below threshold). We discard the attributes of photo in null class because of no obvious attributes in

[1] http://www.flickr.com

this section. The AADB database is split into 8,500 photos for training, 500 photos for validation and 1,000 photos for testing.



(a) (b)

Figure 5: Distributions of the aesthetic scores on the AADB (left) and AVA (right) databases

Table 2: 11 attributes and the number of associated photos on the AADB database. Note that the $\surd$ denotes attributes belong to photo-based attributes or photography-based attributes.

ID	Attribute name	Description	Number of photos	photo-based	photography-based
1	Balancing elements	whether the photo contains balanced elements	10000	$\surd$	
2	Color harmony	whether the overall color of the photo is harmonious	10000	$\surd$	
3	Content	whether the photo has good/interesting content	10000	$\surd$	
4	Depth of field (DoF)	whether the photo has shallow depth of field	10000	$\surd$	
5	Light	whether the photo has good/interesting lighting	10000	$\surd$	
6	Motion blur	whether the photo has motion blur	10000	$\surd$	
7	Object	whether the photo emphasizes foreground objects	10000	$\surd$	
8	Repetition	whether the photo has repetitive patterns	10000	$\times$	$\times$
9	Rule of Thirds	whether the photography follows rule of thirds	10000	$\surd$	
10	Symmetry	whether the photo has symmetric patterns	10000	$\times$	$\times$
11	Vivid Color	whether the photo has vivid color	10000	$\surd$	

The AVA database contains 14 style attributes and these attributes are binary converted to discrete values of 0 or 1. Moreover, this database has about 250,000 photos collected from a social network [2]. Specifically, each photo is randomly assigned with 78 to 549 aesthetic scores ranging from 1 to 10. The AVA database

[2] http://www.dpchallenge.com

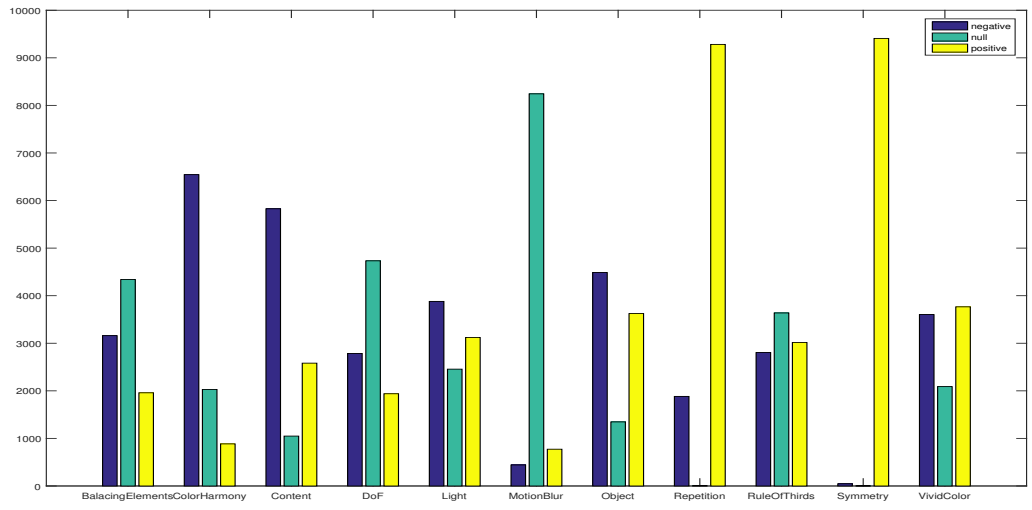


Figure 6: Distributions of the attributes on the AADB databases

Table 3: 14 attributes and the number of associated photos on the AVA database. Note that the $\surd$ denotes attributes belong to photo-based attributes or photography-based attributes.

ID	Attribute name	Description	number of photos	photo-based	photography-based
1	Complementary	whether the photo has pairs of colors	949	$\surd$	
2	Duotones	whether the photo has Duotones	1301		$\surd$
3	HDR	whether the camera reproduce a high dynamic range	396		$\surd$
4	Photo Grain	whether the photo has photo grain	840	$\times$	$\times$
5	Light on White	whether the photo has light on white	1199	$\times$	$\times$
6	Long Exposure	whether the camera use a long-duration shutter speed	845		$\surd$
7	Macro	whether the camera use Macro	1698		$\surd$
8	Motion Blur	whether the photo has motion blur	609	$\surd$	
9	Negative Photo	whether the photo is a additionally color-reversed	959		$\surd$
10	Rule of Thirds	whether the photography follows rule of thirds	1031	$\surd$	
11	Shallow DoF	whether the photo has shallow depth of field	710	$\surd$	
12	Silhouettes	whether the photo match the outline of the objects	1389		$\surd$
13	Soft Focus	whether the lens forms blurred photos	1479		$\surd$
14	Vanishing Point	whether the photo parallel lines to converge	674		$\surd$

is split into training (230,000) and testing (20,000) sets. Since there is no validation set on this database, we select randomly the 20,000 photos from the training set as the validation set. In this database, 14 attributes are provided, including photo-based attributes and photography-based attributes. There are complementary, duotones, HDR, photo grain, light on white, long exposure, macro, motion blur, negative photo, rule of third shallow dof, silhouettes, soft focus and vanishing point. The details of the style attributes and the number of associated photos are listed in Table 3. In these attributes, since the attribute of "Photo Grain" and "Light on White" have not obvious relationship with aesthetics quality, we discard this attribute. The aesthetic scores of photos in this AVA database are showed in Fig.5(b). We find that the aesthetic scores of AVA follow the Gaussian distribution.

6.2. Data Preprocessing and Parameters Settings

We preprocess photos and its attributes provided in two databases. First, we rescale every photo so that the shorter side is of length 256. Then, a 224×224 patch is cropped randomly from the rescaled photo for the purpose of data augmentation [44]. Secondly, on AADB database, the value of attributes in null class is 0. Therefore, we discard these attributes with null class. Then, the aesthetic attributes normalized from the interval of [-1, 1] to [0, 1]. Afterwards, we divide the attributes on AADB whose values are in the interval of (0.5, 1] into high attributes and [0, 0.5) into low attributes. On AVA database, the aesthetic attributes are binary with 0 or 1. Therefore, we define 0 as low attributes and 1 as high attributes. Lastly, the aesthetic scores on two databases are normalized to the interval of [0, 1]. Specifically, we set the threshold C as 0.5 and the score correlation measured by Spearman's ρ between the estimated aesthetic score. The ground-truth scores is employed as performance metrics as in [1].

Due to the fact that most of photos in our experiments are related to natural scenes in daily life, it is helpful to extract the feature representations by the deep models trained on the ImageNet database. We use PyTorch to implement our method and utilize deep residual network (ResNet) [43] model. In order to train the network, we first extract feature representations from the pre-trained ResNet-152 and the size of the feature representation is 2048D. Then, we build the three hidden fully connected layers with ReLU activations. The sizes of these three layers are 512, 128 and 1 respectively. To speed up the convergence rate, in the output layer, we use sigmoid activation to map the aesthetic score into [0,1].

For AADB database, since every photo has their own aesthetic attributes, we can train regression network, pairwise training network and attribute model network in every training sets. However, on the AVA database, only a small portion

of the photos are tagged with aesthetic attributes. Thus, for the photos containing aesthetic attributes, we use regression network, pairwise training network and attribute model, while the photos without aesthetics attributes, we just use the regression network and pairwise training network, setting the attribute model as 0 loss.

6.3. Experimental Results and Analysis

To further demonstrate the effectiveness of our method, we conduct the following four experiments in the two databases. The first one is regression network for aesthetic assessment. Photo aesthetics rating network used in our architecture is fine-tuned from ResNet which we apply Regression loss in our model. The second one is combining regression network and pairwise training network for aesthetic assessment. Photo aesthetics rating network used in our architecture is fine-tuned from ResNet in which we employ the regression loss, ranking loss and Kong[1]'s ranking loss used in our framework for comparison. The third one is model combining regression loss and privileged information loss for aesthetic assessment. The last one is network employing regression loss, ranking loss and privileged information loss.

Table 5 shows the aesthetic assessment results on AADB database and AVA database. From Table 5, we observe as follow:

First, the proposed PI-DCNN method using three loss functions achieves the best performance among all methods with the highest Spearman coefficient ρ. Specifically, compared with photos aesthetic assessment ignoring all privileged information but containing regression network and ranking loss, the proposed method achieves 0.024 and 0.0936 increment of Spearman coefficient, with respect to AADB database and AVA database. Since the method ignoring attributes as privileged information is totally data-driven method, which only learns the mapping from the extracted features to the predictions, ignoring the objective attributes knowledge. The proposed method models privileged information as constraints during model train process, and achieves the best performance in aesthetic assessment.

Second, the ranking+ regression methods have better performance than that only use regression network. Specifically, on AADB database and AVA database, compared with the method just using Regression loss as constraint, the methods utilizing ranking information achieves 0.0081 and 0.0432 increment of Spearman coefficient respectively. Due to the fact that the method ignoring ranking information as constraint only makes the predicted aesthetic score approaching to real

score. Hence, the proposed method utilizing ranking information have better performance.

Third, the privileged information + regression model has a better performance than only utilizing regression loss. To be exact, on AADB database and AVA database, adding the privileged information to the ResNet model can enhance performance by 0.0157 and 0.1154 of spearman coefficient respectively. It is further indicated that the proposed method successfully captures privileged information as constraints during training, exploring both privileged information and training data to obtain better deep network and therefore achieves better performance. Furthermore, compared with ranking + regression methods, the performance of privileged information + regression method has a significant increase. That is mainly because the ranking method and privileged information method have different influence on deep network from different aspects.

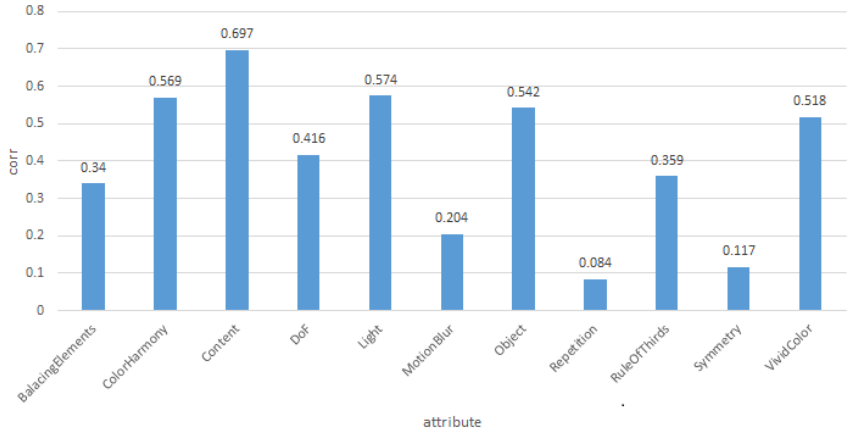


Figure 7: Correlations between attributes and aesthetic score on the AADB databases

6.4. Analysis of Hyper Parameter

As discussed in Eq.11, the hyper parameter C_1 controls the proportion of the ranking loss function. We set the value of C_1 ranging from $\{0.01, 0.1, 0.5, 1, 2, 5\}$ for simplicity and C_1=0.1 achieves the best performance. In addition, the hyper parameter C_2 controls the proportion of the privileged information loss. On the AADB database, we first compute the Pearson's correlation coefficient between the value of attributes and aesthetic score, as shown in Fig.7 The Pearson's correlation coefficient we compute are [0.340, 0.569, 0.697, 0.416, 0.574, 0.204, 0.542, 0.359, 0.518] respectively among these 9 kinds of attributes. This denotes the

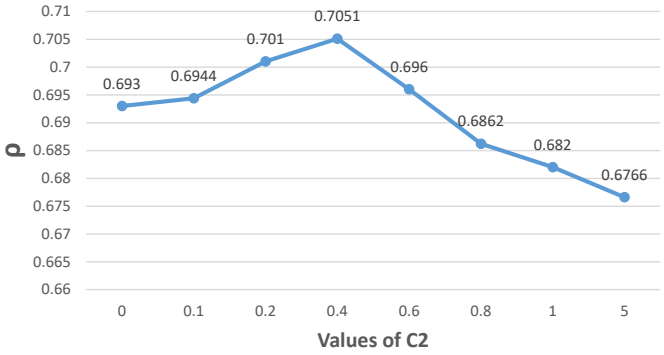


Figure 8: The performance of our model with respect to hyper parameter C_2 on the AADB database

different weights between attributes and aesthetic score. Hence, theoretically, we can set them as the coefficient of 9 privileged information loss in Eq.12 and an optimal value of C_2 exists for the best trade-off between the other loss function and privileged information loss. Then, we conduct experiments with different values of C_2 to explore the the performance of our model. The experimental performance of our proposed PI-DCNN method with respect to C_2 is shown in Fig. 8. As we can see, the performance of Spearman's coefficient is a gradually rise from 0 to 0.4 where it peaks at 0.7051. Afterwards, the performance degenerates rapidly. Actually, compared the method setting the coefficient of all attributes as 1 and adjusting C_2 shown in Eq.13, the result of Spearman's coefficient is 0.6996, lower than Eq.12's result. Hence, The method setting ratio among 9 privileged information loss has better performance.

Table 4: Results on AVA database with each privileged information

PI	PI-DCNN	PI	PI-DCNN
Complementary	0.5721	Negative Photo	0.5606
Duotones	0.5545	Rule of third	0.5437
HDR	0.5317	Shallow DoF	0.5468
Long Exposure	0.5682	Silhouettes	0.5296
Macro	0.5543	Soft Focus	0.5329
Motion Blur	0.5409	Vanishing Point	0.5281

Since the value of attributes are binary with 0 or 1 on AVA database and the Pearsons correlation coefficients cannot be computed, we design an adaptive weighting strategy on AVA database. On AVA database, there are 12 kinds

of attributes belonging to photo-based attributes or photography-based attributes. Hence, we need to optimize 12 weights α^k, k from 1 to 12. We use the values of C_1 and C_2 obtained in AADB experiments and adopt the stochastic gradient descent(SGD) to optimize α^k. Specifically, we separately design each privileged information applied in model to achieve the best result shown in Table 4. Then we initialize α^k with these values in order to reduce training time. Finally, the 12 weights we get are [0.682, 0.540, 0.510, 0.620, 0.473, 0.425, 0.664, 0.432, 0.466, 0.374, 0.421, 0.383] respectively shown in Eq. 14 achieving the best Spearman metric ρ (0.6578).

$$loss_{pi} = C_2\{0.340 * loss_{att1} + 0.569 * loss_{att2} + ... \quad +0.518 * loss_{att9}\} \quad (12)$$

$$loss_{pi} = C_2\{1 * loss_{att1} + 1 * loss_{att2} + ... \quad +1 * loss_{att9}\} \quad (13)$$

$$loss_{pi} = C_2\{0.682 * loss_{att1} + 0.540 * loss_{att2} + ... \quad +0.383 * loss_{att12}\} \quad (14)$$

6.5. Comparison with Related Works

Table 5: Comparison results of photo aesthetic assessment on AADB and AVA database

Model	AADB	Model	AVA
	ρ		ρ
RAPC[1]	0.6782	RAPC[1]	0.5581
Kong's ranking (AlexNet)[1]	0.6515	Kong's ranking (AlexNet)[1]	0.5126
Regression+Kong's ranking (ResNet)[1]	0.6753	Regression+Kong's ranking (ResNet)[1]	0.5430
Square-EMD[45]	0.6889	NIMA (InceptionNet)[48]	0.6120
DCNN (ResNet)[13]	0.6890	DARN[49]	0.5160
Adversarial-DCRN (ResNet)[46]	0.7041	Adversarial-DCRN (ResNet)[46]	0.6313
Multi-task Deep Learning [47]	0.6800		
Regression network (ResNet)	0.6730	Regression network (ResNet)	0.5210
Regression+ranking network (ResNet)	0.6811	Regression+ranking network (ResNet)	0.5642
Regression+PI network (ResNet)	0.6887	Regression+PI network (ResNet)	0.6364
PI-DCNN (ResNet)	**0.7051**	**PI-DCNN (ResNet)**	**0.6578**

To further evaluate the performance of the proposed PI-DCNN method, we choose the following several popular methods for comparison. On the AADB database, we compared our method with the following six works.

[1]In our experiment, Kong's ranking is implemented in ResNet.

- Reg+Rank+Att+Cont (RRAC) model incorporates joint learning of meaningful photographic attributes and image content information to regularize the complicated photo aesthetics rating problem [1].

- Squared Earth Movers Distance (Square-EMD) [45] utilizes the predicted probabilities of all classes and penalizes the miss-predictions according to a ground distance matrix. This method leverages the relationships between classes by training deep nets with the exact squared Earth Movers Distance.

- Attribute-aware Deep Convolution Neural Network (DCNN) [13] jointly learns eight aesthetic attributes along with the overall aesthetic score to perform automatic photo aesthetic assessment.

- Adversarial Deep Convolutional Rating Network (Adversarial-DCRN) [46] introduces a discriminator to distinguish the predicted attributes and aesthetics of the deep network from the ground truth label distribution. Through adversarial learning, the attributes are explored to enforce the distribution of the predicted attributes and aesthetics to converge to the ground truth label distribution.

- Multi-task Deep Learning [47] first learns image aesthetics and personality traits. Then the personality features are employed to modulate the aesthetics features, producing the optimal generic image aesthetics scores to finish aesthetic assessment.

On AVA database, besides RAPC method[1] and adversarial-DCRN[46] method mentioned in AADB, we also consider another two methods for comparison.

- Neural Image Assessment (NIMA) model [48] effectively predicts the distribution of human aesthetic scores using a convolutional neural network, leading to a more accurate quality prediction with higher correlation to the ground truth aesthetic scores.

- Deep Attractiveness Rank Net (DARN) [49] combines rank net trained with a large set of side-by-side multi-labeled image pairs, and deep convolutional neural network to directly learn an attractiveness score mean.

Table 5 shows the experimental results of comparison methods on AADB and AVA database. Spearman metric ρ of PI-DCNN is higher than RRAC, Square-EMD, DCNN, adversarial-DCRN, Multi-task Deep Leaning, NIMA and DARN

methods. RRAC utilizes a pairwise ranking loss to explicitly exploit relative rankings of photo pairs and only one attribute of content is predicted by a added branch of original network. The proposed model (PI-DCNN) improves the ranking algorithm explained in Section5.2 and considers the knowledge underlying more attributes as privileged information without additional branch network, to leverage the relations between privileged information and photo aesthetics. Although Square-EMD model demonstrated its loss function is superior to some works, Square-EMD model is intrinsically a single-task can not handle the information of aesthetic attributes as privileged information. The proposed PI-DCNN model outperforms Square-EMD model, verifying the benefits of using the aesthetic attributes as privileged information. DCNN and Adversarial-DCRN methods learned only a few aesthetic attributes and aesthetic scores independently. Both of them failed in thoroughly exploring the intrinsic relation between privileged information and photo aesthetics. Multi-task deep learning, NIMA and DARN ignore the importance of attributes as privileged information in photo aesthetic assessment, resulting in worse performance than our method.

7. Conclusion

In this paper, we propose a novel method for photo aesthetic assessment through exploring photo attributes as privileged information. A deep convolutional neural network with three types of loss functions in fully connected layer is adopted to learn the mapping from photos to the aesthetic scores. In addition, we utilize the privileged information to propose the probabilistic dependencies and transfer such probabilistic dependencies to the objective function constraints for photo aesthetic assessment. The experimental results on the AADB database and AVA database demonstrate that our privileged information-aware deep network outperforms the state-of-the-art approaches for photo aesthetic assessment.

8. Conflict of interest statement

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work. There is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of the manuscript entitled.

References

[1] S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, Photo aesthetics ranking network with attributes and content adaptation, in: European Conference on Computer Vision, Springer, 2016, pp. 662–679.

[2] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, J. Luo, Aesthetics and emotions in images, IEEE Signal Processing Magazine 28 (5) (2011) 94–115.

[3] Y. Kao, C. Wang, K. Huang, Visual aesthetic quality assessment with a regression model, in: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 1583–1587.

[4] Y. Deng, C. C. Loy, X. Tang, Image aesthetic assessment: An experimental survey, IEEE Signal Processing Magazine 34 (4) (2017) 80–106.

[5] M. Freeman, The complete guide to light & lighting in digital photography, Sterling Publishing Company, Inc., 2007.

[6] J. Itten, Design and form: The basic course at the Bauhaus and later, John Wiley & Sons, 1975.

[7] C. Li, A. C. Loui, T. Chen, Towards aesthetics: A photo quality assessment and photo selection system, in: Proceedings of the 18th ACM international conference on Multimedia, ACM, 2010, pp. 827–830.

[8] W. Luo, X. Wang, X. Tang, Content-based photo quality assessment, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 2206–2213.

[9] Y. Kao, R. He, K. Huang, Deep aesthetic quality assessment with semantic information, IEEE Transactions on Image Processing 26 (3) (2017) 1482–1495.

[10] C. Cui, H. Liu, T. Lian, L. Nie, L. Zhu, Y. Yin, Distribution-oriented aesthetics assessment with semantic-aware hybrid network, IEEE Transactions on Multimedia 21 (5) (2018) 1209–1220.

[11] M. Nishiyama, T. Okabe, I. Sato, Y. Sato, Aesthetic quality classification of photographs based on color harmony, in: CVPR 2011, IEEE, 2011, pp. 33–40.

[12] D. Liu, R. Puri, N. Kamath, S. Bhattachary, Composition-aware image aesthetics assessment, arXiv preprint arXiv:1907.10801 (2019).

[13] G. Malu, R. S. Bapi, B. Indurkhya, Learning photography aesthetics with deep cnns, arXiv preprint arXiv:1707.03981 (2017).

[14] Z. Wang, S. Chang, F. Dolcos, D. Beck, D. Liu, T. S. Huang, Brain-inspired deep networks for image aesthetics assessment, arXiv preprint arXiv:1601.04155 (2016).

[15] X. Lu, Z. Lin, H. Jin, J. Yang, J. Z. Wang, Rapid: Rating pictorial aesthetics using deep learning, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM, 2014, pp. 457–466.

[16] P. Lv, M. Wang, Y. Xu, Z. Peng, J. Sun, S. Su, B. Zhou, M. Xu, Usar: an interactive user-specific aesthetic ranking framework for images, arXiv preprint arXiv:1805.01091 (2018).

[17] V. Hosu, B. Goldlucke, D. Saupe, Effective aesthetics prediction with multi-level spatially pooled features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9375–9383.

[18] R. Datta, D. Joshi, J. Li, J. Z. Wang, Studying aesthetics in photographic images using a computational approach, in: European Conference on Computer Vision, Springer, 2006, pp. 288–301.

[19] V. Vapnik, A. Vashist, A new learning paradigm: Learning using privileged information, Neural networks 22 (5-6) (2009) 544–557.

[20] N. Murray, L. Marchesotti, F. Perronnin, Ava: A large-scale database for aesthetic visual analysis, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2408–2415.

[21] Y. Ke, X. Tang, F. Jing, The design of high-level features for photo quality assessment, in: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 1, IEEE, 2006, pp. 419–426.

[22] Y. Luo, X. Tang, Photo and video quality evaluation: Focusing on the subject, in: European Conference on Computer Vision, Springer, 2008, pp. 386–399.

[23] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2) (2004) 91–110.

[24] X. Zhang, X. Gao, W. Lu, L. He, A gated peripheral-foveal convolutional neural network for unified image aesthetic prediction, IEEE Transactions on Multimedia (2019).

[25] L. Mai, H. Jin, F. Liu, Composition-preserving deep photo aesthetics assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 497–506.

[26] X. Xu, W. Li, D. Xu, Distance metric learning using privileged information for face verification and person re-identification, IEEE transactions on neural networks and learning systems 26 (12) (2015) 3150–3162.

[27] N. Sarafianos, C. Nikou, I. A. Kakadiaris, Predicting privileged information for height estimation, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 3115–3120.

[28] J. Lambert, O. Sener, S. Savarese, Deep learning under privileged information using heteroscedastic dropout, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8886–8895.

[29] S. Dhar, V. Ordonez, T. L. Berg, High level describable attributes for predicting aesthetics and interestingness, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 1657–1664.

[30] T. O. Aydın, A. Smolic, M. Gross, Automated aesthetic analysis of photographic images, IEEE transactions on visualization and computer graphics 21 (1) (2015) 31–42.

[31] M. Fan, R. Huang, W. Feng, J. Sun, Image blur classification and blur usefulness assessment, in: Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on, IEEE, 2017, pp. 531–536.

[32] S. Herron, Technology of duotone color transformations in a color managed workflow, in: Color Imaging VIII: Processing, Hardcopy, and Applications, Vol. 5008, International Society for Optics and Photonics, 2003, pp. 365–371.

[33] H. Wakabayashi, K. Kazami, T. Sosa, H. Miyamoto, Camera having soft focus filter, uS Patent 4,937,609 (Jun. 26 1990).

[34] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, M. Ouhyoung, Personalized photograph ranking and selection system, in: Proceedings of the 18th ACM international conference on Multimedia, ACM, 2010, pp. 211–220.

[35] K. E. Burchett, Color harmony, Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur 27 (1) (2002) 28–31.

[36] L.-C. Ou, M. R. Luo, A colour harmony model for two-colour combinations, Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur 31 (3) (2006) 191–204.

[37] M. A. Robertson, S. Borman, R. L. Stevenson, Estimation-theoretic approach to dynamic range enhancement using multiple exposures, Journal of Electronic Imaging 12 (2) (2003) 219–229.

[38] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, K. Myszkowski, High dynamic range imaging: acquisition, display, and image-based lighting, Morgan Kaufmann, 2010.

[39] T. Clark, Digital Macro and Close-Up Photography For Dummies, John Wiley & Sons, 2011.

[40] T. Padova, K. L. Murdock, Adobe Creative Suite 3 Bible, Vol. 532, John Wiley & Sons, 2008.

[41] P. Beardsley, D. Murray, Camera calibration using vanishing points, in: BMVC92, Springer, 1992, pp. 416–425.

[42] X. Lu, Z. Lin, X. Shen, R. Mech, J. Z. Wang, Deep multi-patch aggregation network for image style, aesthetics, and quality estimation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 990–998.

[43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[44] S. Ma, J. Liu, C. Wen Chen, A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4535–4544.

[45] L. Hou, C.-P. Yu, D. Samaras, Squared earth mover's distance-based loss for training deep neural networks, arXiv preprint arXiv:1611.05916 (2016).

[46] B. Pan, S. Wang, Q. Jiang, Image aesthetic assessment assisted by attributes through adversarial learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 679–686.

[47] L. Li, H. Zhu, S. Zhao, G. Ding, H. Jiang, A. Tan, Personality driven multi-task learning for image aesthetic assessment, in: 2019 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2019, pp. 430–435.

[48] H. Talebi, P. Milanfar, Nima: Neural image assessment, IEEE Transactions on Image Processing 27 (8) (2018) 3998–4011.

[49] N. Ma, A. Volkov, A. Livshits, P. Pietrusinski, H. Hu, M. Bolin, An universal image attractiveness ranking framework, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 657–665.

Yangyang Shu received his B.S. degree in computer science from Anhui University in 2015, and received his M.S. degree in Computer Science in the University of Science and Technology of China in 2018. Now he is currently pursuing his ph.D degree in Engineering and Information Technology in the University of Technology Sydney, Australia. His research interests cover Machine Learning, Affective Computing and Computer Vision etc..

Qian Li received her doctorate from Chinese Academy of Science. She is currently a postdoctoral research fellow at University of Technology Sydney. She is interested in optimization algorithms for machine learning, topological data analysis and statistical causal analysis. She has published several papers in high-impact machine learning conferences such as AAAI, CVPR, WWW, CIKM, ICCS, IJCNN and journals include KAIS, JNCA, JNSM etc. .

Shaowu Liu received his PhD degree in Computer Science from Deakin University in 2016. Currently, he is a postdoctoral research fellow in School of Computer Science and Advanced Analytics Institute, University of Technology Sydney. His current research interests include User Behavior Analytics, Interpretable Machine Learning, and Representation Learning of Knowledge Graphs. His research has been applied to a number of real-world projects in FinTech and Digital Health. Since 2012, he has published 30+ journal and conference articles. For the research community, he has served as co-chair of ES 2016, IIP 2016, KSEM 2017, and KSEM 2019 conferences.

Guandong Xu is Full Professor and Program Leader at School of Software and Advanced Analytics Institute, University of Technology Sydney and he received PhD degree in Computer Science from Victoria University, Australia. His research inter-

ests cover Data Science, Data Analytics, Recommender Systems, Web Mining, User Modelling, NLP, Social Network Analysis, and Social Media Mining. He has published three monographs in Springer and CRC press, and 180+ journal and conference papers including TOIS, TIST, TNNLS, TSC, TIFS, IEEE-IS, Inf. Sci., KAIS, WWWJ, KBS, Neurocomputing, ESWA, Inf. Retr., IJCAI, AAAI, WWW, ICDM, ICDE, CIKM. He is the assistant Editor-in-Chief of World Wide Web Journal and Software, World Wide Web Journal, Multimedia Tools and Applications, and Online Information Review.