Historical Context-based Style Classification of Painting Images via Label Distribution Learning

Jufeng Yang[1], Liyi Chen[1], Le Zhang[2], Xiaoxiao Sun[1], Dongyu She[1], Shao-Ping Lu[1], Ming-Ming Cheng[1]
[1]College of Computer Science, Nankai University, Tianjin, China
[2]Advanced Digital Sciences Center, Illinois at Singapore

ABSTRACT

Analyzing and categorizing the style of visual art images, especially paintings, is gaining popularity owing to its importance in understanding and appreciating the art. The evolution of painting style is both continuous, in a sense that new styles may inherit, develop or even mutate from their predecessors and multi-modal because of various issues such as the visual appearance, the birthplace, the origin time and the art movement. Motivated by this peculiarity, we introduce a novel knowledge distilling strategy to assist visual feature learning in the convolutional neural network for painting style classification. More specifically, a multi-factor distribution is employed as soft-labels to distill complementary information with visual input, which extracts from different historical context via label distribution learning. The proposed method is well-encapsulated in a multi-task learning framework which allows end-to-end training. We demonstrate the superiority of the proposed method over the state-of-the-art approaches on Painting91, OilPainting, and Pandora datasets.

KEYWORDS

painting style classification; label distribution learning; art history

ACM Reference Format:
Jufeng Yang, Liyi Chen, Le Zhang, Xiaoxiao Sun, Dongyu She, Shao-Ping Lu, Ming-Ming Cheng. 2018. Historical Context-based Style Classification of Painting Images via Label Distribution Learning. In 2018 ACM Multimedia Conference (MM '18), October 22-26, 2018, Seoul, Republic of Korea. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3240508.3240593

1 INTRODUCTION

Painting style classification is an attractive topic that can help the public to decode art paintings better and understand the theme and emotion therein [32]. It is a complex cognitive task because multiple visual areas in the human brain are involved in this process [17, 42, 45]. And most art historians agree with the assumption that art can be classified into the specific style [2, 7, 27, 39, 49].

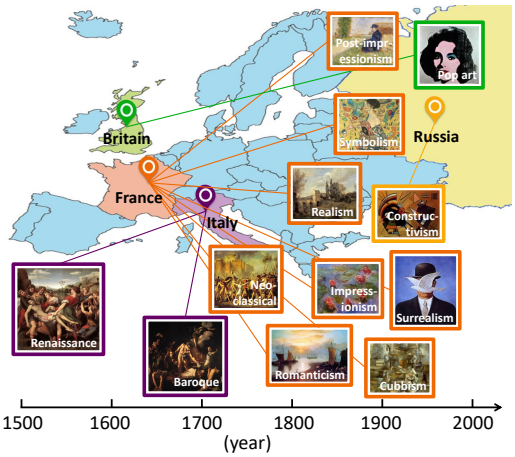

Figure 1: The origin time and birthplace of painting styles in the Painting91 dataset. There are some paintings from different styles and we arrange them according to the chronological order, as indicated by the horizontal axis.

Moreover, it admits high intra-class variations caused by differences between individual creative habit [24]. Since the pioneering work [41] was proposed, several visual features [6, 9, 19, 26, 30, 31] have consistently improved the classification performance leading the research community to address challenging scenarios in complex datasets [3, 5, 23].

Recently, the convolutional neural network (CNN) with hierarchical feature learning capability has led to a breakthrough in many computer vision tasks. Motivated by this, CNN has been widely used to extract the image features for the style classification [1, 33, 34] as well. However, it is widely accepted that the development of painting style is a continuous process and new styles may inherit, develop or even mutate from their predecessors. Furthermore, in actual artistic appreciation, experts usually consider the historical background, politics, religion and other factors when analyzing the style of a painting [18, 46]. From this point of view, painting style evolution is multi-modal because that different input modalities can play an important role in, such as the visual appearance, the birthplace, the origin time. However, existing deep learning methods for painting style classification never utilize multiple input sources mentioned above which encode complementary materials to commonly used visual descriptors.

As shown in Figure 1, twelve paintings from different styles are listed on the axis, and their locations based on their origin time are also marked. We use four marks with different colors to represent

MM '18 , October 22–26, 2018, Seoul, Republic of Korea

ACM ISBN 978-1-4503-5665-7/18/10…$15.00
https://doi.org/10.1145/3240508.3240593

the birthplace of these styles in the map, in which the orange, purple, blue, yellow countries represent Italy, France, Britain, and Russia, respectively. The outer border color of the image corresponds to the color of the marker on the map. We can observe that in some cases the transition of styles is subtle, indicating that the styles produced in the same period often have a high degree of similarity, and for some styles originating from different birthplaces there is usually an obvious gap. Therefore, we argue that a good style classification system should consider different input modalities such as the historical context of the paintings mentioned here as they usually encode complementary information with visual descriptors.

In order to use these historical contexts, we propose to synthesize historical knowledge into the image label via the label distribution learning (LDL) [4, 11–15, 40] which is further employed to generate a proper label distribution in each modality. Multiple label distributions are finally encapsulated into our learning framework which can significantly assist visual feature learning in CNN thus leading to largely improved classification performance. We first study the effect of the style distribution under two kinds of time distributions, a place distribution, and a distribution based on the art movement. Furthermore, we show distributions from each domain are well-complementary and finally a fusion of multiple distributions is advantageous. In the training phase of our experiment, art historical context information (origin time, birthplace, and art movement), which can be easily collected based on the available style label, are converted into a label distribution to assist visual feature learning in CNN. The label distribution works as a strong regularizer by providing a soft label for CNN. During testing, our method works exactly the same way as conventional methods by just a simple feed-forward of only visual inputs images. Extensive experiments show that our method outperforms the state-of-the-art approaches on the Painting91 style dataset, OilPainting dataset, and Pandora dataset. In addition, the proposed method is generic and can be widely applicable for other tasks in which incorporating relevant side-information into visual classification can be beneficial.

Our contributions are summarized as follows. First, we present the historical context-based side information, such as the origin time, birthplace, and the art movement, which may encode complementary information with commonly used visual descriptors. Second, we propose an art historical label distribution learning strategy to encode different input modality into a concrete label distribution. Third, we utilize the knowledge of historical context information to produce a soft label for each painting image and achieve improved performance on three painting datasets. We will make the source code and data available to the public.

2 RELATED WORK

2.1 Style Classification

Painting style classification has been a classical problem in multimedia community and existing methods typically employ various visual descriptors [1, 9, 10, 23, 41]. Sablatnig *et al.* [41] propose a classification system for portrait miniatures and examine the structural signature based on brush strokes, which enables a semi-automatic classification through three different levels of information: the color, shape of the region, and structure of brush strokes. Zujovic *et al.* [56] present some features that address the salient aspects of a painting (*e.g.*, color, texture and edges) in order to classify paintings into genres. A framework integrating multiple visual features is proposed to support automatic classification on large western painting collections [43]. Moreover, Khan *et al.* [23] investigate the effect of various local features and global features (such as bag-of-words framework, LBP, GIST, PHOG, Color GIST, *etc.*) in artist and style classification tasks, and prove combining multiple features can significantly improve the classification performance.

Motivated by the recent success of deep networks on visual recognition, deep learning for art painting classification is also gaining its popularity [20, 28, 29, 47]. Chu *et al.* [5] use the Gram matrix to calculate the correlations between the responses of features extracted by CNN in style classification tasks. The cross-layer features from CNN also present good experimental results on style recognition tasks [33]. Karayev *et al.* [22] define several different types of image style based on several different aspects of visual style, including photographic techniques, composition styles, moods, genres, and types of scenes. For these versatile visual elements, they evaluate single-feature performance (*e.g.*, color histogram, GIST, deep CNN features, *etc.*) as well as second-stage fusion of multiple features. The multi-scale CNN proposed by Peng [34] introduces a method to exponentially generate more training examples with the assumption of label-inheritable property, which can extract multi-scale features from the original and generated images. Another work based on multi-scale features [1] automatically extracts the region of interests (ROI). It describes both holistic and region-of-interests using multi-scale dense convolutional features and separately encodes the two kinds of features using Fisher vector for computational painting categorization.

2.2 Label Distribution Learning

From the statistic point of view, classification is essentially building a mapping from the instances to their labels. Hence, conventional classification approaches typically answer *"which label can describe the instance?"* [12], while little is known about *" how well can this ground truth label describe the instance? "*. Considering our painting style classification here, a painting image often relates to multiple basic styles (e.g., Impressionism, Baroque, Cubism, Expressionism and so on). Each basic style can be correlated and play an important role in the image. For instance, the Baroque style is usually considered evolved from the Renaissance style [8]. The various intensities of all the basic styles naturally form an style distribution which can be captured from various aspects, such as historical context, for the painting images. By regarding the style with the highest intensity or style with higher intensities than a threshold as the positive label, the problem can be naturally solved by existing classification methods. Unfortunately, these approaches will lose the important information of the different intensities of the related styles.

To this end, LDL is proposed to use a certain number of continuous labels to describe one instance. This is usually achieved by defining a probability for each label to represent the degree of how it describes the instance. LDL has established its effectiveness in various vision problems [36, 50, 52, 54, 55]. Yang *et al.* [53] propose the deep age distribution learning (DADL) method to deal with the situation where apparent age estimation differs from chronological age estimation, probably due to the ambiguity from multiple

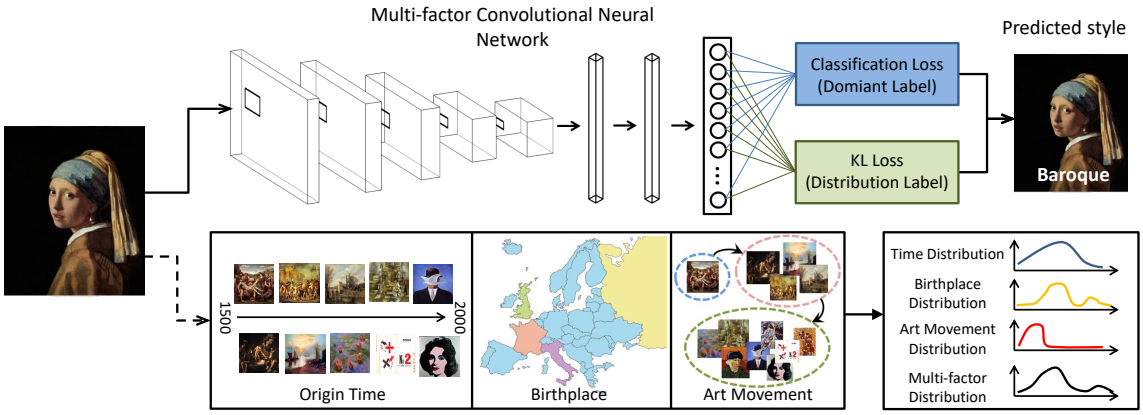

Figure 2: The illustration of the proposed method. Taking into account the three factors in the historical context that describe the relationship between styles (Origin Time, Birthplace, Art Movement), the framework simultaneously optimizes the classification loss and distribution loss. The softmax loss is employed as the classification loss, while the style distribution loss (KL loss) is used as an auxiliary task to assist visual feature learning towards better generalization ability.

individual labelers. In addition, Gao *et al.* [11] demonstrate the superiority of LDL in some domains, such as apparent age estimation, head pose estimation, multi-label classification, and semantic segmentation. Yang *et al.* [51] leverage the ambiguity and relationship between emotional categories to generate emotional distribution, and develop a multi-task deep framework that jointly optimizes classification as well as distribution prediction. While we take inspirations from these methods, we are the first to employ LDL for style classification with our novel enhancement. More specifically, we design proper strategies to generate good label distribution for each historical context. After identifying different label distributions are complementary, multiple label distributions are finally encapsulated into our learning framework which can significantly assist visual feature learning in CNN.

3 METHOD

We propose to generate label distributions considering the impact of three factors which capture complementary information with commonly used visual input in the historical context. Label distributions from three input modalities are well encapsulated into a single distribution finally and serve as a soft-label to enhance the visual feature learning in CNN within a multi-task learning framework. The pipeline of our method is shown in Figure 2.

3.1 Generating Label Distribution

Given painting images $\{x^{(j)}\}_{j=1}^{N}$, we note the ground truth style as $\{y^{(j)}\}_{j=1}^{N}$, where $y^{(j)} \in \{1, \cdots, c\}$. For each instance, we assign a finite set of labels $\{\boldsymbol{l}_i\}_{i=1}^{c}$ representing the degree to which styles describe the image regarding to historical information, where c denotes the number of styles. Note that $\sum_{i=1}^{c} \boldsymbol{l}_i = 1$ and $\boldsymbol{l}_i \in [0, 1]$. In addition, we further propose three strategies to generate the label distributions denoted as $\boldsymbol{t}$, $\boldsymbol{b}$, $\boldsymbol{a}$, which consider the origin time, birthplace, and art movement, respectively.

Time distribution. Motivated by existing single label-based style classification approaches [5, 38], we first generate the label distribution $\boldsymbol{t}$ according to the origin time of the style. Previous attempts [21] turn out to be confined because they overemphasize temporally adjacent styles. To this end, we employ two strategies to generate the time distributions for art paintings, as illustrated in Figure 3. The first time distribution (TD1) represents the style distribution according to the numerical order of the styles. In contrast to TD1 which only considers the sequence of occurrence, the second time distribution (TD2) uses the real origin time of each style. It takes the origin time of style as a condition for generating label distribution.

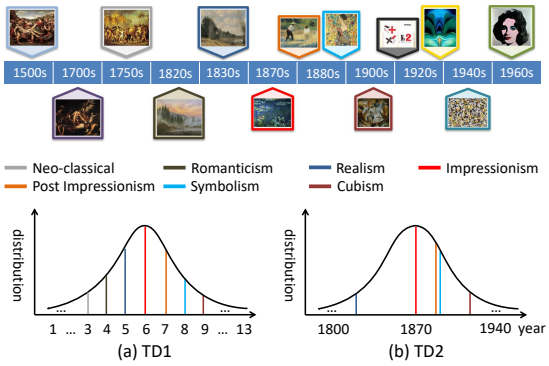

Figure 3: The origin time of 13 styles in the Painting91 dataset and two kinds of time distributions. The 'TD1' represents the label distribution according to the numerical order of the styles, and the horizontal distance between each adjacent style is set as 1. The 'TD2' calculates the label distribution based on the origin time of each style, and the horizontal distance between each adjacent style is normalized according to the origin time of each style. According to 'TD1' and 'TD2', the label distribution of the Impressionism style is generated, and the corresponding styles are also given (best in color view).

Following the Gaussian distribution assumption in LDL works [11, 12, 16], we adopt the typical Gaussian function to generate the time distribution $\boldsymbol{t1}$ for TD1. Each element of $\boldsymbol{t1}$ is denoted as:

$$\boldsymbol{t1}_i = \frac{f(T_i, T_y, \sigma)}{\sum_{k=1}^{c} f(T_k, T_y, \sigma)}. \tag{1}$$

And the probability density function can be written as follows:

$$f(T_i, T_y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{|T_i - T_y|^2}{2\sigma^2}) + \frac{\varepsilon}{c}, \tag{2}$$

where T_i represents the numerical time index of desired style i, and the parameter T_y represents the numerical time index of the ground truth style y. The σ denotes the compactness degree of style in terms of the origin time. Here we fix the parameter ε to 0.1, ensuring that all the styles are taken into consideration with various probabilities. Then we normalize the label to make sure the sum of distribution is 1 as shown in Equation 1.

For TD2, the origin time of each style does not share a unified step size, as illustrated in Figure 3 (b). More specifically, in some time period, there might be only one painting style, while in another period the styles are much more diversified. We thus normalize such origin time of styles using logarithmic calculation before generating the label distribution:

$$\widehat{T}_i = \ln \frac{T_i - T_0}{\delta}, \tag{3}$$

where $\widehat{T}_i$ represents the normalized point-in-time value of desired style i where $\widehat{T}_i$ denotes the time after normalization. The time parameter T_0 as well as the parameter δ are used to narrow the time interval between neighboring styles to facilitate the subsequent learning processing. And we denote each element of time distribution $\boldsymbol{t2}$ as:

$$\boldsymbol{t2}_i = \frac{f(\widehat{T}_i, \widehat{T}_y, \sigma)}{\sum_{k=1}^{c} f(\widehat{T}_k, \widehat{T}_y, \sigma)}, \tag{4}$$

where the parameter $\widehat{T}_y$ represents the normalized point-in-time value of the ground truth style y.

Birthplace distribution. We use the birthplace of style to define the label distribution. Observing the birthplace of various styles, one can see that almost all styles were usually originated from several limited countries, and it is easy to discover the transfer of the world art center. Therefore, we define each label of the birthplace distribution $\boldsymbol{b}$ in the following manner:

$$\boldsymbol{b}_i = \begin{cases} 1, & i = y \\ \frac{\beta}{n_b}, & B_i = B_y, i \neq y \\ 0, & otherwise \end{cases}, \tag{5}$$

where B_i is the birthplace of style i, B_y is the birthplace of ground truth style y, and β controls the weight of the birthplace correlation. If $i = y$, the desired style i is the real style, so it is assigned the value of 1. For other styles that were born in the same country, we note the total number as n_b, and give them the same label value $\frac{\beta}{n_b}$, otherwise zero. The rationale here is that styles originate from the same birthplace should have strong correlations than different styles which generate from different birthplaces. Similarly, we normalize the distribution in Equation 5 to make sure their sum is 1.

Art movement distribution. During a restricted period of time, a group of artists share a specific artistic philosophy or goal, and gradually form a technical tendency or style in art, which is understood as the art movement [38]. In this study, we also investigate the feasibility of encoding the art movement into another label distribution. In our view, each art style can usually be associated with to an art movement period. Puthenputhussery *et al.* [38] show that the art styles belonging to the same art movement period have higher similarity than others that are across different art movement periods. In addition, some styles are considered to have an explicit inheritance relationship with others. Similar to the birthplace distribution, we define the value of ground truth style equals to 1. And for other styles that belong to the same art movements, we give the label with a fixed value, otherwise assigned zero. We define each element of the art movement distribution $\boldsymbol{a}$ as:

$$\boldsymbol{a}_i = \begin{cases} 1, & i = y \\ \frac{\alpha}{n_a}, & A_i = A_y, i \neq y \\ 0, & otherwise \end{cases}, \tag{6}$$

where A_i is the art movement of style i, A_y is the art movement of ground truth style y, and α controls the weight of the art movement correlation. If $i = y$, the desired style i is the real style, so it is assigned the value of 1. For other styles that belong to the same art movement, we note the total number as n_a and give them the same label value $\frac{\alpha}{n_a}$, otherwise assigned zero. In the same way, the distribution in Equation 6 is normalized to ensure that the sum is 1.

3.2 Multi-factor Distribution

As each label distribution aims to capture different art historical context, they usually are well-complementary. This further motivates us to investigate a proper strategy to fuse multiple label distributions which are elaborated in this section.

As described in Figure 3, we have already investigated two different time distribution strategies, i.e. TD1 and TD2, according to the numerical order of occurrence and the point-in-time of origin. While using TD2 can take into account the particularity of style, we also hope to make use of the order of styles (TD1) to make up for the problem of the extremely distant distance between adjacent styles with partial time spans caused by using TD2 alone. In our implementation, we choose TD1 as the main part of the overall time distribution and adjust the distance of some styles that have a close correlation based on TD2. We further consider incorporating the birthplace factor and art movement factor into our final distribution label. In order to consider multiple art historical factors, we integrate all the information with sum pooling operation, which is simple but efficient. And after normalization, the integrated multi-factor distribution is denoted as $\boldsymbol{l}$:

$$\boldsymbol{l} = \eta \times \boldsymbol{t1} + (1 - \eta) \times \boldsymbol{t2} + \boldsymbol{b} + \boldsymbol{a}, \tag{7}$$

where the parameter η controls the influence degree of the point-in-time of style generation. Finally, we normalize $\boldsymbol{l}$ so that $\sum_{i=1}^{c} \boldsymbol{l}_i = 1$.

3.3 Optimization

We assume that we have access to N training samples, and for each sample $x^{(j)}$, we generate the label distribution $\boldsymbol{l}^{(j)}$ using the ground truth label $y^{(j)}$. Our loss function is denoted as follows:

$$L = \lambda L_{sty}(x, y) + (1 - \lambda) L_{dis}(x, \boldsymbol{l}), \tag{8}$$

where the L represents final loss, L_{sty} and L_{dis} respectively denote the classification loss and style distribution loss (KL loss), and the $\lambda \in [0, 1]$ is a weight factor to control the proportion of two losses.

For the classification loss, we calculate the loss of the ground truth and predicted style, and minimize the *softmax function* to optimize it, denoted as:

$$L_{sty}(x, y) = -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{c} \mathbf{1}(y^{(j)} = i) \ln p_i^{(j)}, \tag{9}$$

where $p_i^{(j)}$ indicates the probability of classifying the j^{th} instance $x^{(j)}$ as the i^{th} style. And $\mathbf{1}$ is an indication function where $\mathbf{1}(\epsilon) = 1$ **i.i.f.** $\epsilon = True$, otherwise $\mathbf{1}(\epsilon) = 0$.

For the distribution learning, We employ the Kullback-Leibler (KL) loss following the work of Gao *et al.* [11] and intend to minimize the following KL divergence:

$$KLdiv = \sum_i \boldsymbol{l}_i^{(j)} \ln \frac{\boldsymbol{l}_i^{(j)}}{p_i^{(j)}} \propto \sum_i -\boldsymbol{l}_i^{(j)} \ln p_i^{(j)}. \tag{10}$$

And the style distribution loss measures the KL divergence, which is defined as follows:

$$L_{dis}(x, \boldsymbol{l}) = -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{c} \boldsymbol{l}_i^{(j)} \ln p_i^{(j)}, \tag{11}$$

In this way, we can employ information of $\boldsymbol{l}$, derived from historical contexts such as origin time, birthplace and art movement, as a soft-label to assist visual feature learning in CNN. Hence, our system can not only just learn from *"which style can describe this specific painting image?"*, but also benefit from utilizing side information about *"How well this style can describe this specific painting image?"*.

4 EXPERIMENT

4.1 Datasets

We evaluate the performance of our method for the painting style classification on challenging painting datasets: Painting91 [23], OilPainting [5], and Pandora dataset [9]. The Painting91 dataset consists of 4,266 paintings of 91 painters, in which 2,338 painting images coming from 50 artists are distributed into 13 style categories. We only use these images that have the style label (2,338 images) to evaluate the classification performance. The training set and test set contain 1,250 and 1,088 images respectively, which is consistent with existing work [23]. The OilPainting dataset collects totally 19,787 oil painting images belonging to 17 image styles. The Pandora dataset contains 7,724 images from 12 art styles. Following existing evaluating protocol [5, 9], five-fold and four-fold cross-validation is conducted on the OilPainting and the Pandora dataset respectively.

4.2 Implementation Details

We build a multimodal CNN framework based on the popular deep model VGGNet [44] with 16 layers, which is initialized using the weights trained for the large-scale image classification task [25].

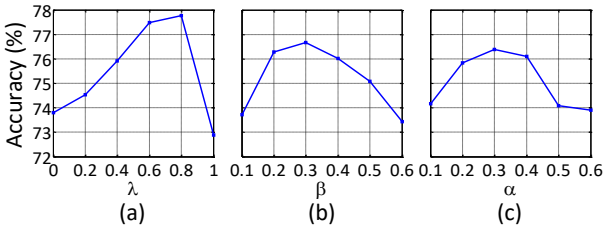

Figure 4: Accuracy performance of the different value of λ, β and α on the Painting91 datasets. We experiment with the value of λ from 0 to 1 in fig(a), and the value of β and α from 0.1 to 0.6 in fig(b) and fig(c).

We change the category number of the *fc8 layer* based on the requirement of datasets because the class number of painting datasets is not equal to ImageNet categories. Different from the original loss layer of VGGNet, we use our multi-task loss including the classification loss and style distribution loss in order to accommodate the style distribution over the three factors we have proposed. Since these three datasets we used only have a single label for each image, firstly we need to generate distribution labels by means of the relationship between various styles, which respectively conform to different factors. And these historical context information are only used to generate the label distribution for the training set.

The learning rate of the last fully-connected layer is initialized 0.01, the batch size of the network is 32, and we fine-tune all layers using stochastic gradient descent. We totally use 6,000 iterations taking about 3h in each experiment, and the learning rate dropped to one-tenth of the original every 1,500 iterations, in order to extract more precise style related information. All our experiments are carried out on NVIDIA GTX TITAN X GPU with 12 GB CPU memory.

4.3 Baseline

For the painting style datasets, most of the existing works recognize the style with the only ground truth label [22, 23, 33]. Based on this, many works achieve promising performance for style classification [1, 5] with VGG network. We apply the probability distribution predicted by the network to compare the generated label distribution with the predefined label distribution, and according to highest probability of style in the distribution to calculate the classification accuracy. In order to verify the effect of the proposed art historical label distribution learning, we choose VGGNet trained with only the ground truth style label as our baseline.

4.4 Distribution Learning Results

We generate the label distributions separately considering the origin time relationship of the style (using the proposed two time distribution strategies), the birthplace relationship and the specific art movement. In addition, we fuse these features of historical context that can comprehensively reflect the relationship of styles in multimodal label distribution experiments. By optimizing the classification loss as well as the style distribution loss, the network learns a variety relationship of historical context between the styles of training.

Table 1: Ablation experiments on the Painting91, OilPainting, and Pandora datasets. The first line denotes baseline using the single label. And we consider four additional properties of historical context with different label distributions. Note that TD1, TD2, BP, and AM represent two time distribution strategies, Birthplace, and Art movement, respectively.

Base	TD1	TD2	BP	AM	Painting91	OilPainting	Pandora
√					72.89%	64.24%	70.52%
	√				76.29%	69.58%	71.09%
		√			75.93%	68.88%	71.12%
			√		76.66%	69.28%	72.21%
				√	76.38%	69.05%	71.95%
	√	√			77.11%	69.85%	71.20%
	√	√	√		77.39%	70.23%	72.87%
	√	√		√	77.21%	70.10%	72.53%
	√	√	√	√	**77.76%**	**70.59%**	**73.28%**

In this section, we first discuss the effect of hyper-parameters in our method, and then we provide ablative experiments to understand the impact using different art historical factors. Finally, we compare the proposed method against the state-of-the-art approaches.

4.4.1 Hyper-parameter. In Figure 4, we show the experiment results using different value of λ, β and α in fig(a), fig(b) and fig(c), respectively. We first research the effect of the value of λ on the experimental results. When the $\lambda = 0.8$, we usually can acquire the best classification accuracy. It means the classification loss is the main part of the final loss because the style classification task is still a single label forecast. So in this paper, we set $\lambda = 0.8$ based on multiple experiments, and we achieve the best classification effect with this value.

We assign the birthplace label value of real style to 1, and for other styles that were born in the same country, we give the fixed value. The final performance is robust to the variations of β in certain ranges. However, setting β to be a tiny value, for example, the label distribution performs similarly to the single label when the $\beta = 0.1$. On the other hand, setting β to a relatively larger value, say $\beta = 0.6$, will over-reduce the effect of the commonly-used ground truth style label, and we observe a downward trend in the final results. In this work, we set $\beta = 0.3$ for birthplace label distribution in order to get the best classification performance. Similarly, we set $\alpha = 0.3$ for the art movement label distribution.

4.4.2 Performance on the Painting91 style dataset. Table 1 shows experimental results using the baseline and our methods using label distribution. We define three different historical context factors (the origin time factors include two representations) that affect the label distribution, where the TD1, TD2, BP, and AM respectively represent the proposed two time distribution strategies, Birthplace, and Art Movement.

The first line of this table shows the results of using the single label, which does not take into account these historical context factors. On the Painting91 style dataset, the classification accuracy based on the single label is 72.89%.

Table 2: Classification performance on the test set of Painting91 dataset, OilPainting dataset, and Pandora dataset. Note that some methods do not provide the source code, thus some datasets cannot be evaluated, denoted as '-'.

Method	Painting91	OilPainting	Pandora
VGGNet [44]	72.89%	64.24%	70.52%
Khan F. S. *et al.* [23]	62.20%	-	-
Condorovici *et al.* [6]	-	-	37.90%
Florea *et al.* [9]	-	-	54.70%
CMFFV [37]	67.43%	-	-
MSCNN1 [34]	69.67%	55.24%	70.32%
MSCNN2 [34]	70.96%	57.92%	69.75%
CNN F4 [33]	69.21%	58.47%	70.47%
Peng K. C. *et al.* [35]	71.05%	-	-
Gram [5]	71.86%	60.61%	-
Gram-Cov [5]	72.41%	60.72%	-
Gram dot Cos [5]	73.59%	63.33%	-
SCMFA [38]	73.16%	-	-
Anwer R. M. *et al.* [1]	74.80%	-	-
Ours	**77.76%**	**70.59%**	**73.28%**

The second to the fifth line are experimental results that define the distribution based on the single historical context factor. The label distribution of 'TD1' strategy achieves a classification accuracy of 76.29% while using the 'TD2' strategy only obtains the result of 75.93%. This is because of the fact that the 'TD2' strategy takes account of the specific production time of each painting style, which usually is not uniform, and the distance between some time adjacent styles is too far in this distribution strategy, so the label value of the corresponding position is very low. This is easy to understand. For example, Renaissance paintings prevailed in the fourteenth to sixteenth centuries, and in the Painting91 style dataset the next style in the chronological order is Baroque that was popular in the seventeenth century. Compared to other adjacent styles that predominated over a few decades or even years, it is too long, even if we take a series of time changes. On contrary, although the 'TD1' only considers the numerical order of style, it can effectively avoid the inequality of style interval time.

The label distribution using birthplace factor has made a greater improvement and achieves 76.66% accuracy, which means the birthplace can effectively reflect the development of styles, and the style produced in the same country under the same historical background is also influenced by other forms of art and therefore shares a high similarity. The classification accuracy using the art movement factor is 76.38%, which demonstrates a definite inheritance relationship in historical context. All the results using single historical context factor can outperform the baseline. In the sixth line, we synthesize two time distributions with $\eta = 0.8$ and achieve better results than using the single time distributions strategy. Compared with using the single factor results, the experiment of using more than one historical context factor can complement each other and achieve better performance. The last line in Table 1 considers multiple historical context factors based on the multimodal CNN framework, and it achieves the best accuracy of 77.76%.

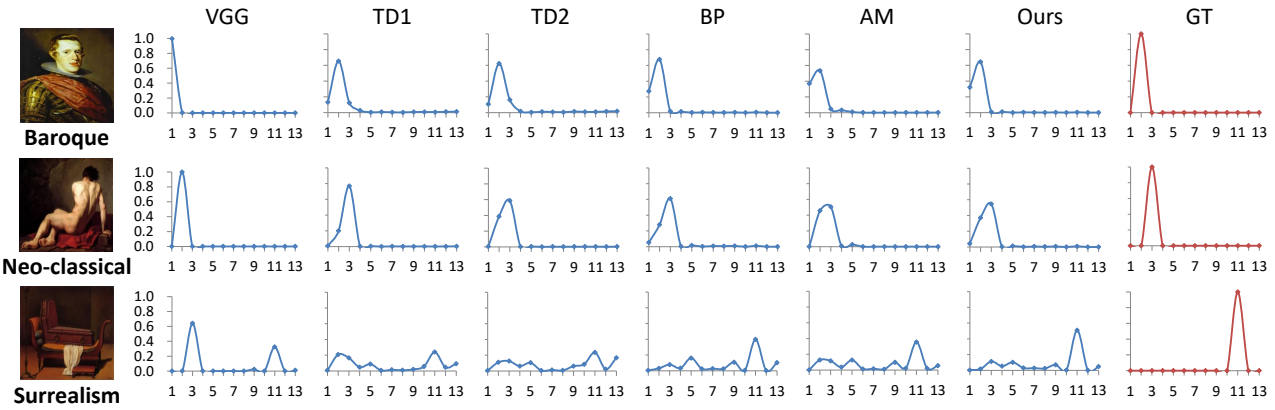

Figure 5: Examples from the Painting91 dataset with the predicted label distribution by VGGNet and our methods. For each subfigure, we introduce the style information at the bottom of the painting. On the right side of the painting, we list six predicted results using single label (VGG) and different label distribution methods (including two time strategies TD1 and TD2, the birthplace distribution (BP), the art movement distribution (AM), and multiple historical context factors (Ours)). The ground truth label (GT) is shown in the last column.

To evaluate the effectiveness of the multi-factor label distribution, we compare our proposed method with deep learning methods and other popular methods on the Painting91 style dataset. Table 2 shows the experimental results of our method as well as other style classification methods for style classification. Khan *et al.* [23] use some common visual feature methods and combine them to get the result of 62.20%. The experimental result using MSCNN [34] is 70.96%, and independently calculating the correlation between Gram matrices and Cosine similarity (Gram dot Cos) [5] achieves a classification accuracy of 73.59%.

In addition, the classification accuracy using SCMFA (a sparse representation based complete kernel marginal Fisher analysis) [38] can achieve 73.16%, and the method of the deep features combining holistic and part-based information [1] is with the best result of 74.80%. In general, our method obtains a classification accuracy of 77.76%, which outperforms the current state-of-the-art methods.

4.4.3 Performance on the OilPainting dataset. The classification result of the single label is 64.24%, without considering the origin time, the birthplace of the style, and the artistic style relationship reflected by art movement. For the single factor distribution, the 'TD1' strategy has achieved a classification accuracy of 69.58%, and the result of 'TD2' strategy is 68.88%, which are similar to the experimental result on the Painting91 dataset. However, different from Painting91, the label distribution experiment using the birthplace factor has only been made a less improvement compared with the single label result. This result reflects that the styles in OilPainting dataset are not very discriminatory for the place of origin. In this dataset, 12 painting styles are originated from France, and the remaining five styles come from other three regions, which means that there is a geographical connection between more than two-thirds of these styles. This would be the main reason why we can not generate a discernible style distribution through the geographic information. As an example, the Renaissance style paintings are divided into three categories: High Renaissance, Northern Renaissance, and Mannerism (Late Renaissance). The Renaissance was born in Italy and later throughout Europe, and the Northern Renaissance refers particularly the Renaissance occurred in Europe north of the Alps. Therefore, it is not appropriate to classify its birthplace as Italy (the birthplace of other two Renaissance styles) or the northern region as a place of origin. The label distribution experiment using the art movement relationship achieves a classification accuracy of 69.05%.

As shown in Table 2, we compared the experimental results of the proposed method with the existing methods on OilPainting dataset, and the classification result of the VGGNet is 64.24%. The classification accuracy using the Gram matrices can be achieved 60.61%, and the method of independently calculating the correlation between Gram matrices and Cosine similarity (Gram dot Cos) achieves the best classification accuracy of 63.33% [5]. Notice that our method obtains a classification accuracy of 70.59%, which outperforms the current state-of-art methods by over 7%.

4.4.4 Performance on the Pandora dataset. The classification result of the baseline is 70.52% without considering the historical context. For the single factor distribution, the 'TD1' and 'TD2' strategies achieve similar results (*i.e.*, 71.09% and 71.12% respectively). Although in this dataset the origin time intervals of some style origins are larger than those in such as OldGreekPottery and Iconoclasm, we still observe some advantages comparing with the baseline. In addition, the label distributions of the birthplace and the art movement factor display the better classification effect, which achieves the accuracy rate of 72.21% and 71.95% separately. It indicates that the birthplace and the art movement information can better reflect the connection between styles in Pandora dataset. Combining multiple historical context knowledge further improves the classification results, and finally, the proposed method achieves the classification accuracy of 73.28% using four historical information.

As shown in Table 2, we compare the multi-factor method with other previous methods on the Pandora dataset. The classification result of traditional visual descriptor method using local and global

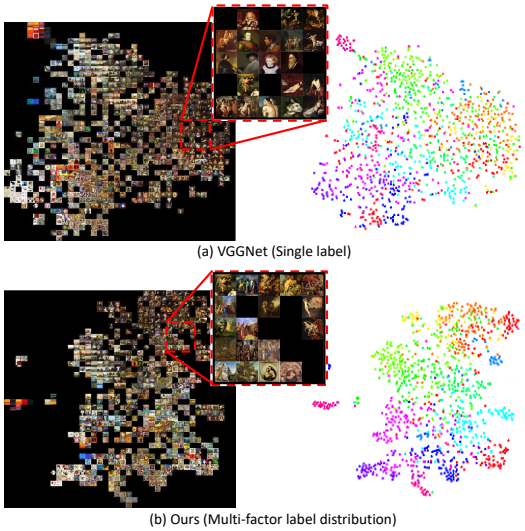

Figure 6: Comparison of VGGNet and our method on the Painting91 dataset. For each subfigure, the left image is two-dimensional representation using t-SNE for test images, and the right is the feature embedding with 13 style categories denoted by different colors. The VGGNet is trained in the discrete label space, while the proposed method takes the style relation into consideration. Please, zoom-in for details of the border between the Baroque and the Renaissance.

features [9] is 54.70%. The classification accuracy of MSCNN1 [34] is 70.32%, and the CNN F4 [33] method achieves 70.47%. The single label method using VGGNet [44] gets the result of 70.52%. Our method considering the multi historical factors outperforms these methods and achieves the accuracy of 73.28%.

4.5 Visualization and Analysis

As shown in the Figure 5, we list some representative painting images from the validation dataset of Painting91, and compare the predicted results between our method and the single label method using VGGNet for each painting. In each group of distribution, the red distribution represents the ground truth label (GT), and blue distributions represent six predicted results using the single label (VGGNet) and different historical context (two time strategies TD1 and TD2, the birthplace distribution (BP), the art movement distribution (AM), and multiple historical context factors (Ours)). In each distribution image, the ordinate indicates the probability of each label, and the abscissa indicates 13 styles in Painting91 dataset, of which the numbers 1 to 13 respectively represent Renaissance, Baroque, Neo-classical, Romanticism, Realism, Impressionism, Post-impressionism, Symbolism, Cubbism, Constructivism, Surrealism, Abstract expressionism, and Pop art.

In Figure 5, the first painting is Baroque style, and the single label method is very firmly that this is a Renaissance style painting. In our methods, both two kinds of time distributions show a decreasing trend on both sides of the ground truth style except for the difference of each distribution value, and predicted distributions are roughly consistent with the generated distributions using historical context. For the birthplace distribution, the second highest is the Renaissance that is also born in Italy. The second painting comes from the Neo-classical style which is often confused with Baroque, probably because they have similar color characteristics and the strong contrast of light and shade. This feature is also reflected in time distributions since the Baroque and the Neo-classical style both originate before the 19th century. For these two paintings, the classification advantage of the art movement distribution is not very obvious (the second highest value approximates the probability of the correct style) due to the confusion of the image itself. But sometimes art movement factor also shows the better differentiation. The last painting belongs to the Surrealism style, while it also has strong baroque and neoclassical features due to its dim shades. The VGGNet misclassified it into Neo-classical style, and our methods using different historical context are also affected by the diverse features of the painting, especially the TD1. Fortunately, the birthplace distribution and art movement distribution can effectively alleviate this situation. Therefore, our method affects style classification via combining many factors in historical contexts that reflect the development of painting style.

As shown in the Figure 6, we compare our method with the single label method in the Painting91 dataset. Two groups of fig (a) and fig (b) show the results of the VGGNet and the proposed label distribution method. We show the two-dimensional representation using t-SNE [48] on the left, and the right image is the feature embedding with 13 style categories denoted by different colors. Figure 6 (a) (left) shows the single label method cannot accurately distinguish the style of painting, especially the three styles of Renaissance, Baroque, and Neoclassical. Figure 6 (b) (left) shows our method can effectively distinguish Baroque and Renaissance style paintings because of the art historical context knowledge. And we can clearly see that the method we proposed can cluster paintings with the same style together compared to the single label method.

5 CONCLUSION

Different from the existing methods that only use visual features for painting image style classification, we show that historical context-based side information may encode complementary information. Motivated by this, a multi-factor distribution is employed, based on an ensemble of label distribution learning with these three factors, as a soft label to enhance the feature discriminability in CNN. We achieve this knowledge distilling through a multi-task learning framework which is end-to-end-trainable. Experimental results demonstrate that our proposed method successfully embodies the relationship of painting styles in historical context and performs favorably against the state-of-the-art approaches on various painting style datasets.

ACKNOWLEDGMENTS

This research was supported by Natural Science Foundation of Tianjin, China (NO. 18JCYBJC15400), the Open Project Program of the National Laboratory of Pattern Recognition (NLPR), NSFC (NO. 61620106008, 61572264), the national youth talent support program, Tianjin Natural Science Foundation for Distinguished Young Scholars (NO. 17JCJQJC43700), Huawei Innovation Research Program.

REFERENCES

[1] Rao Muhammad Anwer, Fahad Shahbaz Khan, Joost van de Weijer, and Jorma Laaksonen. 2016. Combining holistic and part-based deep representations for computational painting categorization. In *ACM International Conference on Multimedia Retrieval*.

[2] Sidney J Blatt and Ethel S Blatt. 2014. *Continuity and change in art: The development of modes of representation.* Routledge.

[3] Gustavo Carneiro, Nuno Pinho da Silva, Alessio Del Bue, and João Paulo Costeira. 2012. Artistic image classification: An analysis on the printart database. In *European Conference on Computer Vision*.

[4] Ke Chen, Joni-Kristian Kämäräinen, and Zhaoxiang Zhang. 2016. Facial age estimation using robust label distribution. In *ACM International Conference on Multimedia*.

[5] Wei-Ta Chu and Yi-Ling Wu. 2016. Deep Correlation Features for Image Style Classification. In *ACM International Conference on Multimedia*.

[6] Razvan George Condorovici, Corneliu Florea, and Constantin Vertan. 2015. Automatically classifying paintings with perceptual inspired descriptors. *Journal of Visual Communication & Image Representation* 26 (2015), 222–230.

[7] Ralph Fanning and Helen Gardner. 1959. Art through the ages. *College Art Journal* 8, 4 (1959), 329.

[8] Lois Fichner-Rathus. 2011. *Foundations of Art and Design: An Enhanced Media Edition.* Cengage Learning.

[9] Corneliu Florea, Răzvan Condorovici, Constantin Vertan, Raluca Butnaru, Laura Florea, and Ruxandra Vrânceanu. Pandora: Description of a painting database for art movement recognition with baselines and perspectives. In *European Signal Processing Conference*.

[10] Corneliu Florea, Cosmin Toca, and Fabian Gieseke. 2017. Artistic movement recognition by boosted fusion of color structure and topographic description. In *IEEE Winter Conference on Applications of Computer Vision*.

[11] Bin-Bin Gao, Chao Xing, Chen-Wei Xie, Jianxin Wu, and Xin Geng. 2017. Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing* 26, 6 (2017), 2825–2838.

[12] Xin Geng. 2016. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering* 28, 7 (2016), 1734–1748.

[13] Xin Geng and Peng Hou. 2015. Pre-release prediction of crowd opinion on movies by label distribution learning. In *International Joint Conference on Artificial Intelligence*.

[14] Xin Geng, Qin Wang, and Yu Xia. 2014. Facial age estimation by adaptive label distribution learning. In *International Conference on Pattern Recognition*.

[15] Xin Geng and Yu Xia. 2014. Head pose estimation based on multivariate label distribution. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[16] Xin Geng, Chao Yin, and Zhi-Hua Zhou. 2013. Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 10 (2013), 2401–2412.

[17] Joseph Goguen. 1999. *Art and the brain.* Imprint Academic.

[18] Ernst Hans Gombrich and EH Gombrich. 1995. *The story of art.* Vol. 12. Phaidon London.

[19] Daniel J Graham, James M Hughes, Helmut Leder, and Daniel N Rockmore. 2012. Statistics, vision, and the analysis of artistic style. *Wiley Interdisciplinary Reviews: Computational Statistics* 4, 2 (2012), 115–123.

[20] Samet Hicsonmez, Nermin Samet, Fadime Sener, and Pinar Duygulu. 2017. DRAW: Deep networks for Recognizing styles of Artists Who illustrate children's books. In *ACM International Conference on Multimedia Retrieval*.

[21] Zengwei Huo, Xu Yang, Chao Xing, Ying Zhou, Peng Hou, Jiaqi Lv, and Xin Geng. 2016. Deep age distribution learning for apparent age estimation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

[22] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. 2014. Recognizing image style. In *British Machine Vision Conference*.

[23] Fahad Shahbaz Khan, Shida Beigpour, Joost Weijer, and Michael Felsberg. 2014. Painting-91: A large scale database for computational painting categorization. *Machine Vision & Applications* 25, 6 (2014), 1385–1397.

[24] Roy King, Jith Meganathan, Jill Nagahara, and Marilia Boscolo. 1998. Individual differences in complexity preference and artistic style: Neoclassical versus expressionistic aesthetics. *Empirical Studies of the Arts* 16, 1 (1998), 15–23.

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*.

[26] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 2 (2004), 91–110.

[27] Thomas Munro. 1946. Style in the arts: a method of stylistic analysis. *The Journal of Aesthetics and Art Criticism* 5, 2 (1946), 128–158.

[28] Loris Nanni and Stefano Ghidoni. 2017. How could a subcellular image, or a painting by Van Gogh, be similar to a great white shark or to a pizza? *Pattern Recognition Letters* 85 (2017), 1–7.

[29] Loris Nanni, Stefano Ghidoni, and Sheryl Brahnam. 2017. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition* 71 (2017), 158–172.

[30] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 7 (2002), 971–987.

[31] Aude Oliva and Antonio Torralba. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42, 3 (2001), 145–175.

[32] DM Parker and Jan B Deregowski. 1991. *Perception and artistic style.* Elsevier.

[33] Kuan-Chuan Peng and Tsuhan Chen. 2015. Cross-layer features in convolutional neural networks for generic classification tasks. In *IEEE International Conference on Image Processing*.

[34] Kuan-Chuan Peng and Tsuhan Chen. 2015. A framework of extracting multi-scale features using multiple convolutional neural networks. In *IEEE International Conference on Multimedia and Expo*.

[35] Kuan-Chuan Peng and Tsuhan Chen. 2016. Toward correlating and solving abstract tasks using convolutional neural networks. In *IEEE Winter Conference on Applications of Computer Vision*.

[36] Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew C Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions.. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[37] Ajit Puthenputhussery, Qingfeng Liu, and Chengjun Liu. 2016. Color multi-fusion fisher vector feature for fine art painting categorization and influence analysis. In *IEEE Winter Conference on Applications of Computer Vision*.

[38] Ajit Puthenputhussery, Qingfeng Liu, and Chengjun Liu. 2016. Sparse representation based complete kernel marginal fisher analysis framework for computational art painting categorization. In *European Conference on Computer Vision*.

[39] Stephanie Ross. 2003. Style in art. *The Oxford handbook of aesthetics* (2003), 228.

[40] Christian Rupprecht, Iro Laina, Robert DiPietro, Maximilian Baust, Federico Tombari, Nassir Navab, and Gregory D Hager. 2017. Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In *IEEE International Conference on Computer Vision*.

[41] Robert Sablatnig, Paul Kammerer, and Ernestine Zolda. 1998. Hierarchical classification of paintings using face- and brush stroke models. In *International Conference on Pattern Recognition*.

[42] Lior Shamir and Jane A Tarakhovsky. 2012. Computer analysis of art. *Journal on Computing & Cultural Heritage* 5, 2 (2012), 1–11.

[43] Jialie Shen. 2009. Stochastic modeling western paintings for effective classification. *Pattern Recognition* 42, 2 (2009), 293–301.

[44] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.

[45] Robert L Solso. 1996. *Cognition and the visual arts.* MIT press.

[46] Hippolyte Taine. 1865. *The philosophy of art.* Baillie re.

[47] Wei Ren Tan, Chee Seng Chan, Hernán E Aguirre, and Kiyoshi Tanaka. 2017. Fuzzy qualitative deep compression network. *Neurocomputing* 251 (2017), 1–15.

[48] Laurens Van Der Maaten. 2014. Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research* 15, 1 (2014), 3221–3245.

[49] Heinrich Wölfflin. 2012. *Principles of art history.* Courier Corporation.

[50] Chao Xing, Xin Geng, and Hui Xue. Logistic boosting regression for label distribution learning. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[51] Jufeng Yang, Dongyu She, and Ming Sun. 2017. Joint image emotion classification and distribution learning via deep convolutional neural network. In *International Joint Conference on Artificial Intelligence*.

[52] Jufeng Yang, Ming Sun, and Xiaoxiao Sun. 2017. Learning visual sentiment distributions via augmented conditional probability neural network. In *AAAI Conference on Artificial Intelligence*.

[53] Xu Yang, Bin Bin Gao, Chao Xing, Zeng Wei Huo, Xiu Shen Wei, Ying Zhou, Jianxin Wu, and Xin Geng. 2015. Deep label distribution learning for apparent age estimation. In *IEEE International Conference on Computer Vision*.

[54] Sicheng Zhao, Hongxun Yao, and Xiaolei Jiang. 2015. Predicting continuous probability distribution of image emotions in valence-arousal space. In *ACM International Conference on Multimedia*.

[55] Ying Zhou, Hui Xue, and Xin Geng. 2015. Emotion distribution recognition from facial expressions. In *ACM International Conference on Multimedia*.

[56] Jana Zujovic, Lisa Gandy, Scott Friedman, Bryan Pardo, and Thrasyvoulos N Pappas. 2009. Classifying paintings by artistic genre: An analysis of features & classifiers. In *IEEE International Workshop on Multimedia Signal Processing*.