Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Photograph aesthetical evaluation and classification with deep convolutional neural networks

Yunlan Tan[a,b,c], Pengjie Tang[b,c], Yimin Zhou[c], Wenlang Luo[a,b,*], Yongping Kang[a,b], Guangyao Li[c]

[a] School of Electronics and Information Engineering, Jinggangshan University, Ji'an, Jiangxi, China
[b] Key laboratory of watershed ecology and geographical environment monitoring, National Administration of Surveying, Mapping and Geoinformation, Ji'an, Jiangxi, China
[c] College of Electronics and Information Engineering, Tongji University, Shanghai, China

ARTICLE INFO

Keywords:
Image aesthetics
Quality assessment
High aesthetics
Low aesthetics
Image classification
Deep convolutional neural network
GoogLeNet
Feature representation

ABSTRACT

In response to the growth of digital photography and its many related applications, researchers have been actively investigating methods for providing automated aesthetical evaluation and classification of photographs. For computational networks to recognize aesthetic qualities, the learning algorithms must be trained using sample sets of characteristics that have known aesthetic values. Traditional methods for developing this training have required manual extraction of aesthetic features for use in the practice datasets. With abundant appearance of convolutional neural networks (CNN), the networks have learned features automatically and have acted as important tools for evaluation and classification. At the time of our research, several existing convolutional neural networks for photograph aesthetical classification only used shallow depth networks, which limit the improvement of performance. In addition, most methods have extracted only one patch as a training sample, such as a down-sized crop from each image. However, a single patch might not represent the entire image accurately, which could cause ambiguity during training. What's more, for existing datasets, the numbers of high quality images of each category are mostly too small to train deep CNN networks. To solve these problems, we introduce a novel photograph aesthetic classifier with a deep and wide CNN for fine-granularity aesthetical quality prediction. First, we download a large number of consumer photographic images from DPChallenge.com (a well-known online photography portal) to construct a dataset suitable for aesthetic quality assessment. Then, we zoom out the images into 256×256 by bilinear interpolation and crop 10 patches (Center+four Corners+Flipping). Once we have associated the set with the image's training labels, we feed the images with the bag of patches into the fine-tuned networks. Our proposed computational method is configured to classify the photographs into high and low aesthetic values. A training pattern specifying an output of (0, 1) indicates that the corresponding image belongs to the "low aesthetic quality" set. Likewise, a training pattern with an output of (1, 0) indicates that the corresponding image belongs to the "high aesthetic quality" set. Experimental results show that the accuracy of classification provided by our method is greater than 87.10%, which is noticeably better than the state-of-the-art methods. In addition, our experiments show that our results are fundamentally consistent with human visual perception and aesthetic judgments.

1. Introduction

With the rapid growth of the imaging and mobile technologies, taking digital photographs has become a daily life activity these days. People want to take, share, and view photographs that have high aesthetic quality. According to statistics provided by DPChallenge.com [1], photographers have uploaded over 587,000 professional photographs. This convenience has stimulated the development of network sharing centers and portals that attract professional or amateur photographers from around the world. Therefore, automatic assessment of aesthetic quality of photographs is a promising technique in many applications related to production, enhancement, management, retrieval, and recommendation of photographs. It is a challenge to establish the standards for differentiating between high and low quality images. First, visual data are very rich and ambiguous, because aesthetical assessment and prediction of the aesthetic value of images are highly subjective and not universal. Second, when judging photographs, people often apply their own cultural relativity to aesthetic judgments. Finally, even if we could gain agreement that certain degradations of photos (e.g., an image that is out of focus) are indicators of poor quality, it is far more difficult to find consensus about higher level positive visual properties such as color harmonies,

* Corresponding author at: School of Electronics and Information Engineering, Jinggangshan University, Ji'an, Jiangxi, China.

http://dx.doi.org/10.1016/j.neucom.2016.08.098
Received 17 February 2016; Received in revised form 16 August 2016; Accepted 31 August 2016
Available online xxxx
0925-2312/

layout, and lighting. With all these challenges, one might question the possibility of creating generic models that encode photographic preferences and then developing an automatic classifier that can learn and use these preferences accurately. Nevertheless, automatic assessment of photo quality from the perspective of visual aesthetics is of great interest in high-level vision research and has drawn much attention in recent decades [2].

Image features (e.g., low-level, high-level) are adopted by computer vision researchers for the purpose of objective image quality assessment [3], and many successful algorithms have been proposed for this purpose [4–8]. Wang and Datta were the first to realize the quantization of image features, including brightness, color distribution, wavelet, region composition and depth of field. By applying a support vector machine (SVM) or linear regression to distinguish high from low quality photographs, they achieved an accuracy of 70.12%. Luo et al. [5] proposed a photo quality assessment method that first extracted the subject region from a photo, and then formulated a number of high level semantic features for photograph quality classification. Ke et al. [6] designed some high level semantic features based on the spatial distribution of edges, blur, and the histograms of low-level color properties such as brightness and hue. They tested their features on a large, diverse dataset, and their system was able to achieve a classification rate of 72%. Tong et al. [7] tried to classify photos as either professional or snapshots, but they used the Corel image database, which is too homogeneous to separate the two classes. In addition, they simply collected a large set of low level features from the image retrieval literature. Liu et al. [8] presented a Multiple Kernel Learning (MKL) method for aesthetic image classification without using explicit feature selection steps. They produced better results than [6,7] using fewer features, but the results still had a classification rate of only 78.3% on a set of 6000 data. Sun et al. [9] tried to design optimized visual features to mimic human perception in photo assessment. Wong et al. [10] presented a saliency-enhanced image classification method, but this method must be preprocessed to detect the salient region. Zhao and Huang et al. [11] proposed a novel local binary count descriptor for rotation invariant texture classification. Guo et al. [12] used both hand-crafting and semantic features to improve the performance of automatic aesthetic assessment of images. Different from reference images, Zhang et al. [13] assessed the image quality with no reference image by using sparse feature representation. They extracted features from image structure patches and PCMs (pixel correlative matrix). Their method achieved comparable high accuracy.

Apart from image features extraction, it is helpful to develop new image enhancement tools to make images look better with photo aesthetic quality assessment [14–18]. Bhattacharya et al. [14] presented an interactive application that enables users to improve the visual aesthetics of their digital photographs by using spatial recomposition. Rather than prescribing a user-guided image segmentation and inpainting solution, Xu et al. [15] developed a photo-taking interface that provides real-time feedback on how to position the subject of interest according to the photography composition rule-of-thirds. Aydin et al. [16] presented a perceptually calibrated system for automatic aesthetic evaluation of photographic images based on a set of fundamental and meaningful aesthetic attributes such as sharpness, depth, clarity, colorfulness, and tone. They can give “aesthetic signature” of each image automatically and various photo editing applications. Wang et al. [17] built an image aesthetic classification and evaluation system with Hadoop cloud computing. Because of using parallel-processing strategy, the system shorts the processing time of images aesthetic analysis and has an efficient aesthetic evaluation for the application of mobile devices.

In addition, researchers have explored the optimization of image processing algorithms using perceptual measures of visual quality as objective functions [19–23]. Marchesotti et al. [19] proposed a method using BoVW (Bag of Visual Words) and FV (Fisher Vector) to encode the feature vectors using descriptors such as SIFT (Scale Invariant Feature Transform). Their method outperforms previous methods by a significant margin. Their method learned the linear SVMs with a hinge loss using the primal formulation and a Stochastic Gradient Descent (SGD) algorithm, but they achieved just 68.55% accuracy. Jiang et al. [20] proposed a novel regression method named Diff-RankBoost based on RankBoost and support vector techniques. They predicted coarse-granularity aesthetic categories for more than 450 real consumer photographic images. Wang et al. [21] performed large scale studies analyzing algorithm performance, namely the structural similarity index (SSIM). They developed the ACQUINE [22] aesthetics value measurement system, an aesthetic evaluation and search engine. However, this method is impractical for large datasets due to the limits of SVM. Wu et al. [23] extended their SVM classification method to predict aesthetic adjectives rather than aesthetic scores. They introduced a probabilistic post-processing step that alleviated effects due to misleadingly labeled training data.

Most previous works on aesthetic image analysis have focused on designing appropriate features. Those methods suffer from several drawbacks. First, all attributes that are related to image aesthetic quality assessment cannot be pointed out. In fact, it is hard to discover all the attributes that can affect image aesthetic quality. Researchers just adopt a small number of them which are well-known and easy to be implemented. Second, it is hard to explain how those image features affect the image aesthetic quality. For example, we may agree on that image color will affect the image aesthetic quality a lot. Professional photographers carefully set the image color to gain better visual effect. But there is no fixed discipline that the color scheme of beautiful images must be followed. Both images with simple color palette and images which are colorful can have high aesthetic quality. Third, it is not easy to describe the image features accurately with mathematical model even we are sure that the features can affect the image aesthetic quality.

Since designing handcrafted features has long been regarded as an appropriate method for predicting image aesthetics, none of the above researchers delved further into the classification methods with regard to machine learning. More recently, new research efforts in deep learning methods have brought breakthroughs in many traditional computer vision problems, becoming one of the most powerful learning architectures for many vision tasks including object recognition, image classification, and video classification [24–28]. CNNs are designed for hierarchical feature representation mechanisms from lower level to higher level in which each level consists of a certain number of feature maps. The feature maps of each level are obtained from the maps contained in the previous level by applying several operations such as linear convolution, non-linear activation, and spatial pooling. Researchers who explored these methods [26–28] were able to show some improvement in automatic aesthetical assessment by modifying the network structures slightly (e.g., adding a layer or adding a column) or by adjusting the training strategy (e.g., fine-tuning). Besides useful techniques such as ReLU, dropout, and data augmentation introduced in [24], there are several efficient CNN approaches to improve performance in various classification problems. Wu et al. [29] modeled a weakly supervised, deep multiple instance learning framework, and they achieved convincing performance in vision tasks including classification and image annotation. To improve the geometric invariance of CNN activations, Gong et al. [30] presented a simple but effective scheme called multi-scale orderless pooling (MOP-CNN). Generally, the existing deep convolutional neural networks (DCNN) require a fixed-size input image. To eliminate the above constraint, He et al. [31] equipped the networks with a spatial pyramid pooling strategy. Their new network structure (SPP-net) could boost the accuracy of image classification, allowing their methods to rank third in image classification in ILSVRC 2014.

Multi-column neural networks [32,33] have been demonstrated to be an efficient approach to improving the performance of single-column neural networks in various classification problems. In multi-