A multi-scene deep learning model for image aesthetic evaluation

Weining Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai, Xiangmin Xu

www.elsevier.com/locate/image

PII: S0923-5965(16)30066-2
DOI: http://dx.doi.org/10.1016/j.image.2016.05.009
Reference: IMAGE15087

To appear in: *Signal Processing : Image Communication*

Received date: 17 October 2015
Revised date: 10 May 2016
Accepted date: 11 May 2016

Cite this article as: Weining Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai and Xiangmin Xu, A multi-scene deep learning model for image aesthetic evaluation, *Signal Processing : Image Communication* http://dx.doi.org/10.1016/j.image.2016.05.009

A Multi-Scene Deep Learning Model for Image Aesthetic Evaluation

Weining Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai, Xiangmin Xu*

School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China

*Corresponding author. Tel: +86 13808848846, Email address: xmxu@scut.edu.cn

Abstract

Aesthetic evaluation of images has attracted a lot of research interests recently. Previous work focused on extracting handcrafted image features or generic image descriptors to build statistical model for aesthetic evaluation. However, the effectiveness of these approaches is limited by researchers' understanding on the aesthetic rules. In this paper, we present a Multi-Scene Deep Learning Model (MSDLM) to enable automatic aesthetic feature learning. This deep learning model achieves better results because it improves performance on some major problems, including limited data amount and categories, scenes dependent evaluation, unbalanced dataset, noise data etc. Major innovations are as follows. (1) We design a scene convolutional layer consist of multi-group descriptors in the network elaborately so that the model has a comprehensive learning capacity for image aesthetic. (2) We design a pre-training procedure to initialize our model. Through pre-training the multi-group descriptors discriminatively, our model can extract specific aesthetic features for various scenes, and reduce the impact of noise data when building the model. Experimental results show that our approach significantly outperforms existing methods on two benchmark datasets.

Keyword: Deep learning, Image aesthetic, Multi-scene deep learning model, Pre-training

1. Introduction

Volume of images grows explosively through social network in past years. It is an arduous task for people to find and manage high quality photos in large amount of collections. Aesthetic analysis can help people to select beautiful images automatically, filter out the unappealing ones, provide aesthetic feedbacks etc. [20]. It also helps people to construct a harmony human-machine interactive system [2]. The aesthetic quality has become an important factor for image retrieval engines [1].

The aesthetic evaluation of images aims at building a computational model to simulate human's aesthetic perception. The evaluation model can give aesthetic scores to images, and classify them into groups, e.g. high quality or low quality.

In early time, researchers extracted image features, and then trained the models by using machine learning methods. In order to get good evaluation results, researchers spent huge effort on designing descriptive features based on aesthetic rules which inspired by domain knowledge from painting, photography, art, the human aesthetic feeling and visual attention mechanism [3,5-6,11-13,21] etc. Tong et al. [11] adopted many low-level features such as color histogram and image energy. Ke et al. [19] designed high-level features e.g. spatial distribution of edges and color distribution, and achieved better results with a much smaller number of features. Datta et al.