Accepted Manuscript

A New Deep Representation for Large-scale Scene Classification

Bo Dai, Feng Mei, Liujing Xu, Deliang Ji, Jia Shi

PII: S1047-3203(18)30192-5
DOI: https://doi.org/10.1016/j.jvcir.2018.08.005
Reference: YJVCI 2254

To appear in: *J. Vis. Commun. Image R.*

Received Date: 24 June 2018
Revised Date: 6 August 2018
Accepted Date: 6 August 2018

Please cite this article as: B. Dai, F. Mei, L. Xu, D. Ji, J. Shi, A New Deep Representation for Large-scale Scene Classification, *J. Vis. Commun. Image R.* (2018), doi: https://doi.org/10.1016/j.jvcir.2018.08.005

A New Deep Representation for Large-scale Scene Classification

Bo Dai[1],Feng Mei[1],Liujing Xu[1], Deliang Ji[2] and Jia Shi[2]
State Grid Zhejiang Electric Power Company Information & Telecommunication Branch, Hangzhou, China.
Zhejiang Huayun electric power industrial group

Abstract: Large scale scene classification based on image is an important problem in computer vision. In this paper, we propose a method to fuse the local features of scene images into a geometric feature that can reflect both the geometric features of the scene image and the color intensity distribution. First, each scene image is segmented into a set of individually connected regions according to their color intensity distribution. A region adjacency graph is constructed to encode the geometric properties and color intensity of scene images. Later, a 5 tier CNN architecture was constructed to study regional features. Then, a thinning process is carried out to obtain a discriminant and compact template set from the training rag. These templates are used to extract graphlets finished bag (r-bogs) images represented by each scene. Finally, the strategy of boosting development is to classify the extracted r-bogs scenes. Experimental results on different datasets demonstrate the effectiveness and effectiveness of the proposed method.

Keywords: scene classification, graphlets, finished bag, region adjacency graph

1. Introduction

Large scale scene classification is an important problem in many computer vision applications, in machine learning and pattern recognition [36]-[38]. However, it is still a successful large-scale scene classification on two factors of the challenges: on the one hand, the component of scene images and their spatial relations are complex and unstable, which makes it difficult to extract features, enough discrimination scene classification; on the other hand, a large number of components and their bilateral the relationship between challenge, the computer can be handled effectively. Therefore, in large-scale scene classification, more discriminative and compact are becoming more and more important.

In the development of image analysis, many features are proposed, which can be divided into two categories: global and local. Global features, such as space [1], represent an image by a single vector, and are therefore tractable for traditional classifiers such as support vector machines (SVM). However, global functional occlusion and clutter are sensitive and lead to low classification accuracy. Contrary to the global features, local features, such as scale invariant feature transform (SIFT) [2], are extracted at interest points, which are robust to image deformation. Different images may produce different amounts of local features. To facilitate the processing of traditional classifications, these local features tend to incorporate feature representations into unordered bags. Unfortunately, as an unstructured representation, the feature description package ignores the spatial relations of the local features, thus avoiding its discrimination.

In order to encode the local features in scene classification, several local feature fusion methods are proposed. Lazebnik et al. [3] presents the spatial Pyramid matching kernel (SPMK), according to its geometric type clustering of image local features, and the integration of different geometric types into the final kernel. However, [4] retains no discrimination information, so it is not the best classification task. In [5], the implied de Lickley distribution (LDA) is used to simulate scene images, that is, each image is represented by a set of codewords generated by the corresponding potential topics. Because LDA ignores the spatial relation of image blocks, an improved LDA (6) is proposed, taking into account the spatial relation of image blocks. However, the generation process of these LDA based methods is time-consuming and difficult to apply to large-scale scene classification. Local feature integration based on graph [7] model the local feature of the image into a graph. In [8], each image is modeled as a tree, and the image matching is constructed as tree matching. Unfortunately, compared with the general graph, the ability to model trees with regional relationships is limited. Keselaman et al. [11] defines a graph called the most common abstraction (LCA) that represents a spatial relation consisting of an object; however, LCA cannot be output to traditional classifications, such as support vector machines [13], directly. Walking kernel [9] uses a finite sequence of adjacent regions to capture local features of the walking structure. Unfortunately, the structural sway phenomenon [10] brings about noise, and therefore limits it to more discriminating.

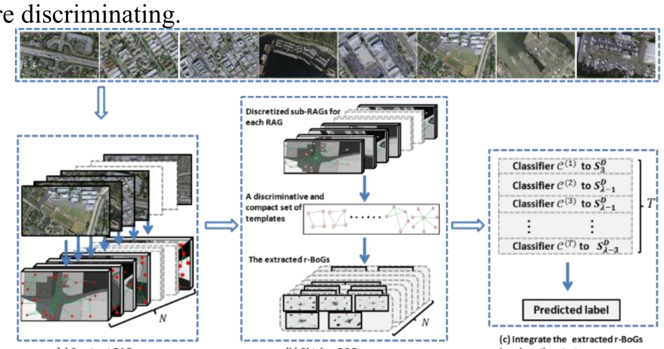

Figure1. The flowchart of our approach

To solve or at least reduce the above problem, a new representation method, r-bogs, is proposed for large-scale scene classification. As shown in Figure 1, each scene image is partitioned into a separate set of connected regions. These single connected region spatial relation model, construct a graph is called a tattered rag, namely, each vertex represent the local characteristics of the distribution of single connected regions and connected edge space of adjacent areas (Figure 1 (a)). Measuring a ragged similarity, it is a direct comparison of all 22 graphlets, that is, a link to the tattered ones, they are. Unfortunately, based on graph theory, the number of graphlets in a block is huge, so the scheme is computationally intractable. In order to achieve an efficient metric, a refinement procedure is carried out to obtain a set of discriminative and compact template sets, i.e., discrete blocks. These templates are used to extract the corresponding r-bogs (Figure 1 (b)). To extract r-