Optik 156 (2018) 791–796

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Semantic Constraint Based Target Object Recognition

Hao Wu [a], Rongfang Bie [a], Junqi Guo [a], Xin Meng [b], Shenling Wang [a,*]

[a] *College of Information Science and Technology, Beijing Normal University, China*
[b] *Electric Power Planning & Engineering Institute, China*

ARTICLE INFO

Article history:
Received 1 November 2017
Accepted 13 December 2017

Keywords:
Semantic constraint
Object recognition
Wordnet subtree
Candidate learning instance joint entropy

ABSTRACT

With the growth of deep learning, object recognition has received increasing interests and its accuracy has been improved significantly in the past few years, However, high-quality recognition largely depends on a large number of learning instances. If the number of learning instances is reduced, it's difficult to maintain realistic recognition accuracy. Moreover, traditional methods usually don't consider the semantic relationship between different regions. Actually, semantic constraint would contribute to improve the recognition accuracy effectively.

Aiming at the problems above, we proposed one semantic constraint based object recognition method. On the one hand, instance-based transfer learning model could make use of learning instances of other categories to maintain realistic recognition accuracy. On the other hand, semantic constraint between different regions simulated as joint entropy is used to recognize target object more accurately. At last, adequate experiments using a large number of images show that our model not only could reduce the number of learning instances but also could achieve realistic recognition.

1. Introduction

Object recognition [1,2] is used to recognize specific object in the image. In the last few years, this technology in the field of computer vision contributed to find and identify objects in an image or video sequence. In the traditional recognition process, some classic feature descriptors, such as SIFT [3], GIST [4] and HOG [5], could extract the features of images effectively. Based on them, many optimized feature descriptors [6–11] could extract the features more fast or adequately. For a long time, SVM combined with optimized feature descriptors are used to estimate the specific category of each object. Moreover, above-mentioned models are also effective for mage classification [12,13], image retrieval [14], and image annotation [15,16]. Within a long duration, recognition methods are on the foundation of improvement for feature extraction and classification. In recent years, with the development of hardware, deep-seated relationship between different pixels could be extracted more adequately, the overwhelming experimental results also verify the performance of deep learning. The essence of deep learning is to extract deep level information through complicated structure and parameters using a large number of learning instances. From all of them, CNN-based methods [17–19], RBM-based methods [20–22], Autoencoder-based methods [23–25] and Sparse coding-based Methods [26–28] are often considered as four classic categories.

As discussed above, quite a few methods have been used to maintain or improve the recognition accuracy. Especially for deep learning based models, they nearly have replaced all traditional methods using overwhelming experimental results.

* Corresponding author.
E-mail address: shenlingwangbnu@163.com (S. Wang).

https://doi.org/10.1016/j.ijleo.2017.12.033

More importantly, further improvements based on deep learning models are still promising. Promising experimental results indeed cover up a lot of shortcomings but shortcomings really exist obviously. Firstly, a large number of learning instances are one non-ignorable burden. If adequate learning instances from website and some other databases are needed, collection process would waste lots of human resource. In some other cases, it's difficult for us to retrieve enough learning instances even if we would like to spend lots of human resource. For instance, even if we collect all red-crowned crane images of websites, they still couldn't fulfill the requirements of deep learning model. Although some previous methods [29–32] have already reduced the number of learning instances through decision model optimization or learning instance replacement, the majority of learning instance reduced models are still too complicated which has become one new burden for human resource and computing resource.

Moreover, the majority of recognition methods ignore the semantic relationship between different regions. Actually, the sematic constraint between different regions could contribute to recognize the target object effectively. For instance, we can't ensure whether the target object is building, but the possibility of building existence is increased through recognizing its neighbor region as street. Even if some methods have taken the semantic relationship between different regions into consideration, the models are still not suitable for complicated images. If they are applied to complicated images concluding more regions, the effectiveness is reduced obviously.

Aiming at the problems above, we combined instance-based transfer learning model and semantic constraint model to achieve realistic recognition using relatively few learning instances. The main contributions of this paper are in the following:

(1) Instance-based transfer learning model is used to reduce the number of learning instances through thought of learning instance substitution.
(2) Semantic constraint based model is used to improve the object recognition accuracy through semantic relationship constraint between different regions simulated as joint entropy.

2. Algorithm

In this process, we mainly presented how to achieve high-quality recognition. Compared to traditional recognition methods, instance-based transfer learning and semantic constraint are essential contributions, then we would introduce them in details.

2.1. Instance-based transfer learning

As is known to us, the quality of deep learning model largely depends on a large number of learning instances. If the number of learning instances is reduced, it's difficult to construct one CNN model, not to mention maintain realistic accuracy. Aiming at this problem, we drew on the experience of instance-based transfer learning idea to reduce the number of learning instances. Based on this idea, the selection of candidate learning instance has become one pending problem. Although some complicated models have been used to select the candidate learning instances effectively, excessive consumption of resources are special burden for us. So in this paper, a relatively simple model of Wordnet subtree [33] is used as reference to select the suitable candidate learning instances.

The similarity of two categories is defined by the number of nodes shared by their parent branches, divided by the length of the longer of the two branches (Fig. 1). The similarity could be defined by:

$$S_{ij} = \text{intersect}(\text{par}(i), \text{par}(j))/\max(\text{length}(\text{par}(i)), \text{length}(\text{par}(j))), \tag{1}$$

For instance, the similarity between "tabby cat" and "felis domesticus" is 0.93, while the similarity between "tractor trailer" and "felis domesticus" is 0.21. We could judge whether the unknown learning instances are candidate learning instances through Wordnet subtree based similarity. In most cases, if the value of Wordnet subtree based similarity is over 0.5, we consider it as a candidate learning instance. Based on the theory of reinforcement learning, we give each candidate learning instance as different enhancement weights referenced by the value of Wordnet subtree based similarity. More concretely, if the candidate learning instances are more semantically similar to target object, we give them more weights. Otherwise, they are given as less weights.

In the process of recognition model construction, similar to structure and parameters in the paper [34], we extract a 4096-dimensional feature vector from the target region using Caffe [35] model. The model is introduced by Krizhevsky and features are computed by forward propagating a mean-subtracted 227*227 RGB image through five convolutional layers and two fully connected layers. The network architecture details could be learnt from papers [35,36].

2.2. Semantic constraint

As discussed above, traditional methods usually recognize the target object without considering the semantic relationship between different regions. Actually, the semantic relationship would contribute significantly to improve the recognition